SENSING THE CITY: URBAN ANALYSIS FOR THE XXI CENTURY

enrique.friasmartinez AT Telefonica.com
www.enriquefrias-martinez.info

6.8 billion subscriptions
96% of world's population (ITU)

Mobile penetration of 120% to 89% of population (ITU

More time spent on our phones than watching TV or
with our partner (US and UK)

Emerging and developed regions

Cell Phones as Sensors of Human Activity

Digital footprints enable large-scale analysis of human behavior

Bits

Business ▪ Innovation ▪ Technology ▪ Society

May 19, 2011, *7:06 pm* **The Sensors Are Coming!**

By NICK BILTON

ieee spectrum INSIDE TECHNOLOGY

Telecom / Wireless

NEWS

Cellphones for Science

Scientists want to put sensors into everyone's hands

Unprecedented Historic Moment

Digital Footprints

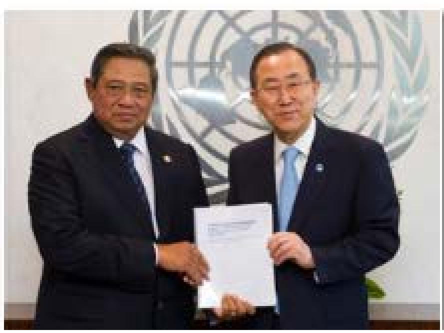
For the first time in human history, we have access to large-scale human behavioral data at varying levels of spatial and temporal granularities

Collecting **large volumes** of **real data** about **human urban behavior** is a challenging yet high value problem

High-level Panel ON
the Post-2015 Development Agenda

HOME | ABOUT | THE PANEL | THE SECRETARIAT | OUTREACH | THE REPORT | MEDIA

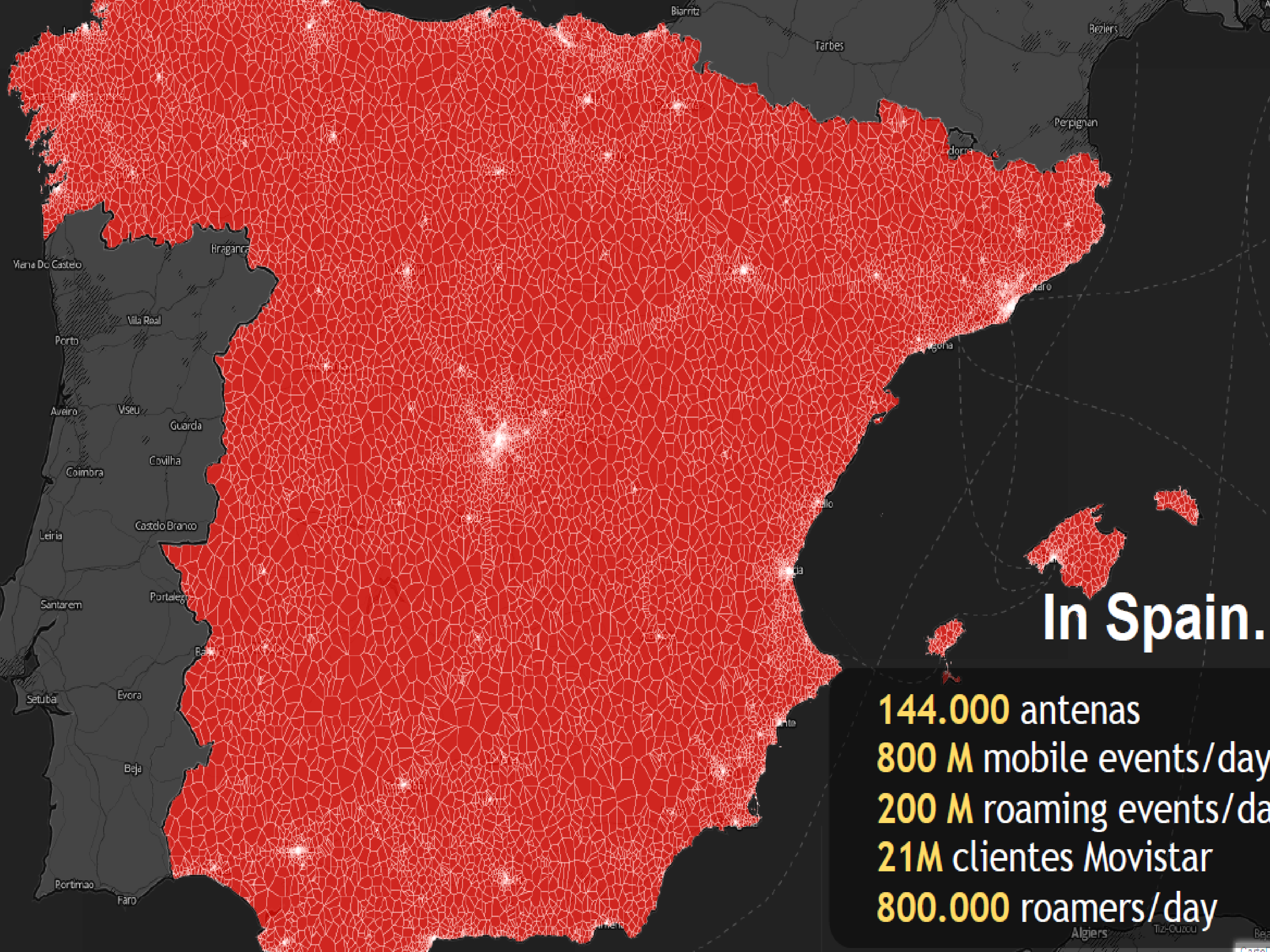
HIGH LEVEL PANEL RELEASES RECOMMENDATIONS FOR WORLD'S NEXT DEVELOPMENT AGENDA



Eminent Persons from Around the World Call for a New Global Partnership to Eradicate Poverty and Transform Economies through Sustainable Development

The High Level Panel on the Post-2015 Development Agenda today released **"A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development,"** a report which sets out a universal agenda to eradicate extreme poverty from the face of the earth by 2030, and deliver on the promise of sustainable development. The report calls upon the world to rally around a new Global Partnership that offers hope and a role to every person in the world.

Wanted: A data revolution


Biarritz
Tarbes
Beziers
Perpignan
Viana Do Castelo
Bragança
Vila Real
Porto
Aveiro
Viseu
Guarda
Covilha
Coimbra
Castelo Branco
Leiria
Santarem
Portalegre
Setubal
Evora
Beja
Portimao
Faro
Algiers
Tizi-Ouzou
In Spain.
144.000 antenas
800 M mobile events/day
200 M roaming events/da
21M clientes Movistar
800.000 roamers/day

Typical Mobile Data

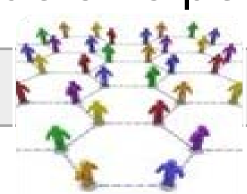
- CDR

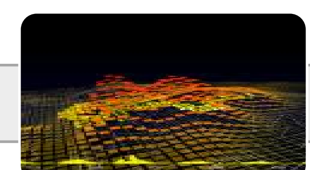
HR_ORG	TLFN_A	TLFN_B	CD_GEO_A	CD_GEO_B	DT_ORG	CD_SNTD	CD_ERB	CD_CCC	QT_DUR
20:05:31	XXX	YYY	3	11	20140519	2	1562	568	33
…	…	…	…	…	…	…	…	…	…

- SMS

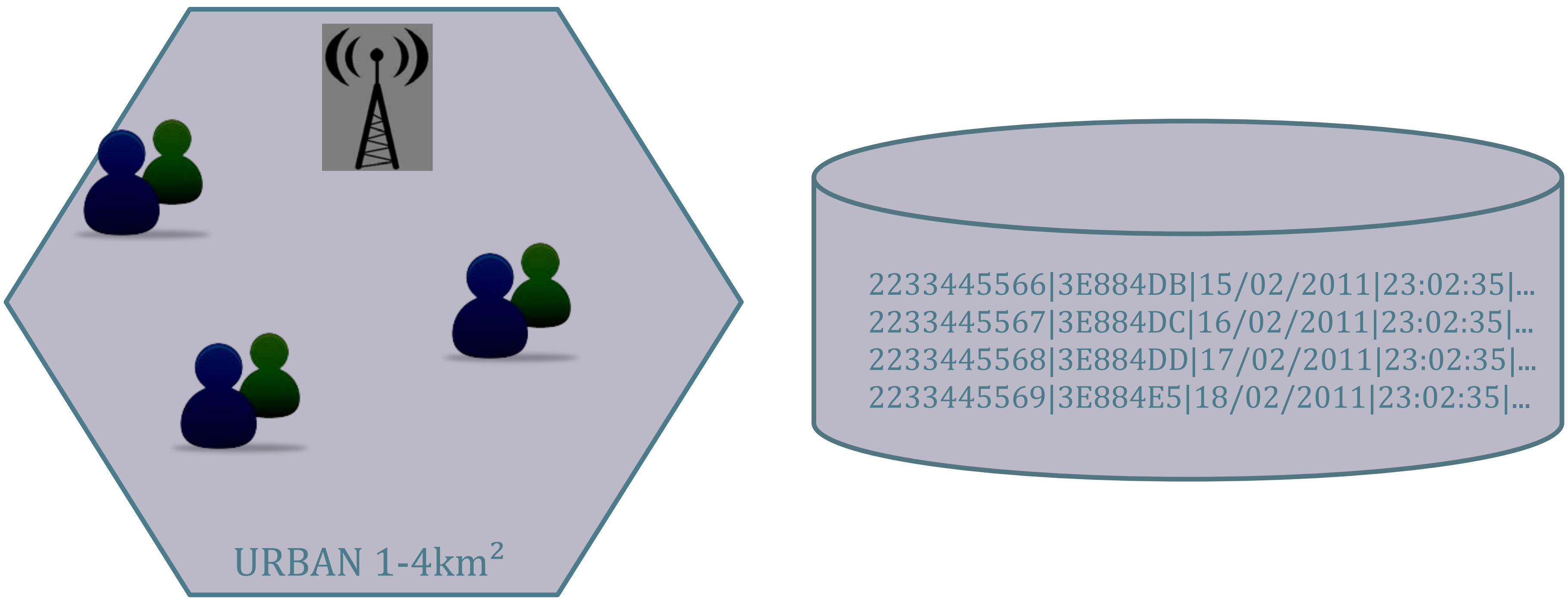
HR_ORG	TLFN_A	TLFN_B	CD_GEO_A	CD_GEO_B	DT_ORG	CD_SNTD	QT_TRFG
15:53:54	XXX	ZZZ	3	25	20140506	2	1
…	…	…	…	…	…	…	…

Consumption	**Social Network**	**Mobility**
Call duration	In/Out Degree	Radius of gyration
N. Events	Delta w.r.t time window	Travelled distance
Lapse between events	Unique Calls per day	Rate of popular antennas
Reciprocated events	Unique SMS per day	Regularity of popular





Call Detail Records

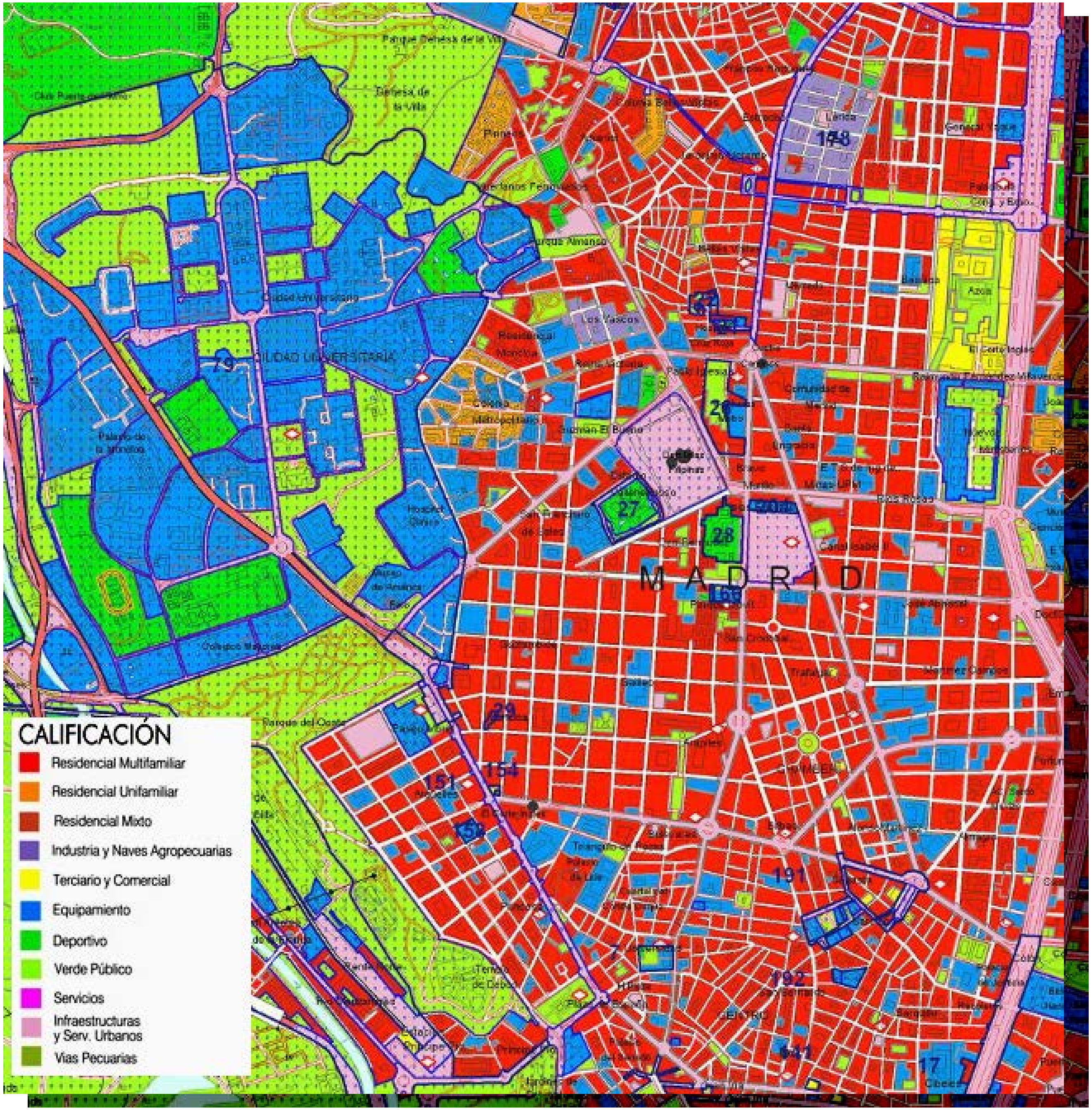


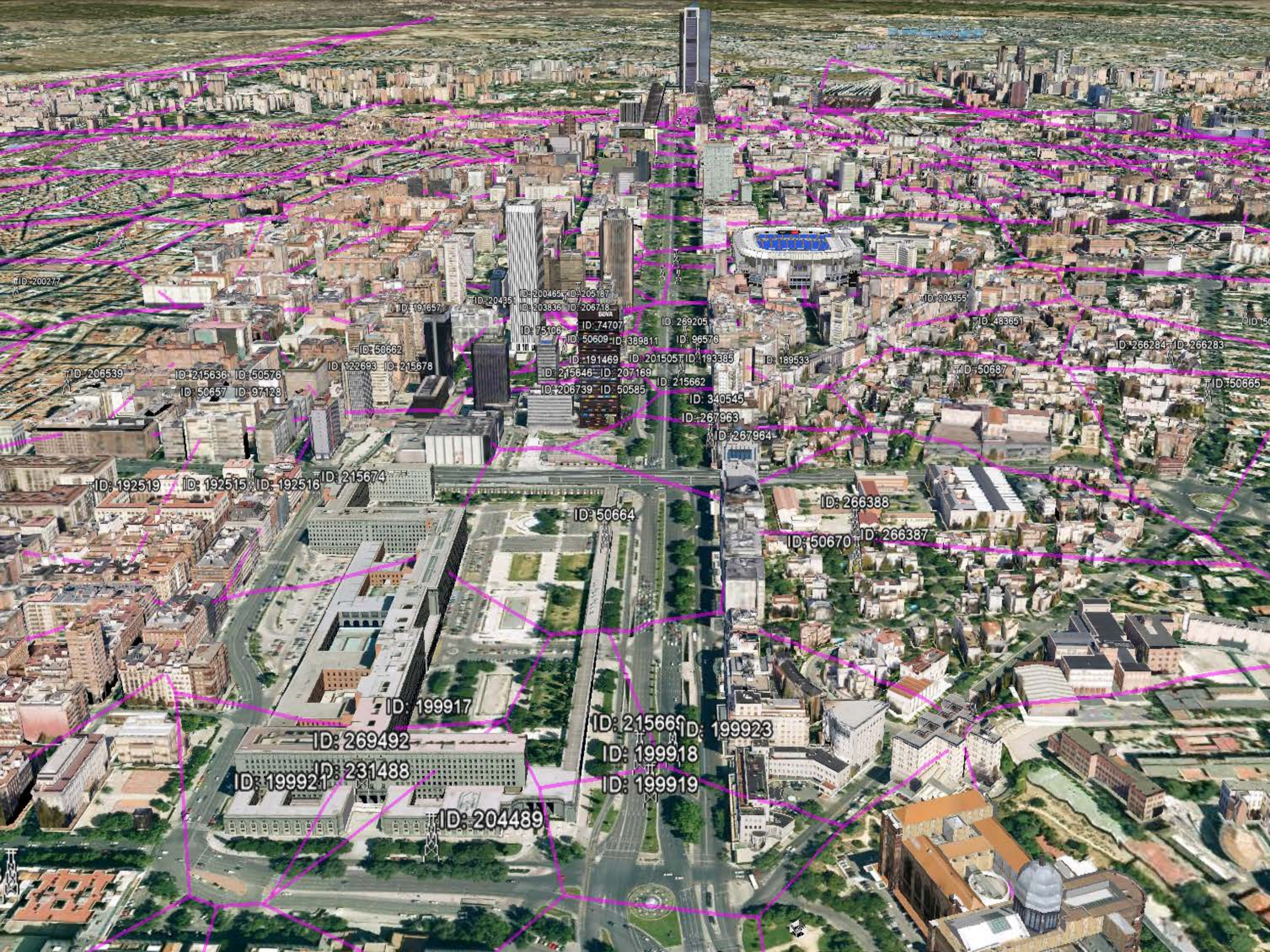
flickr

facebook

foursquare

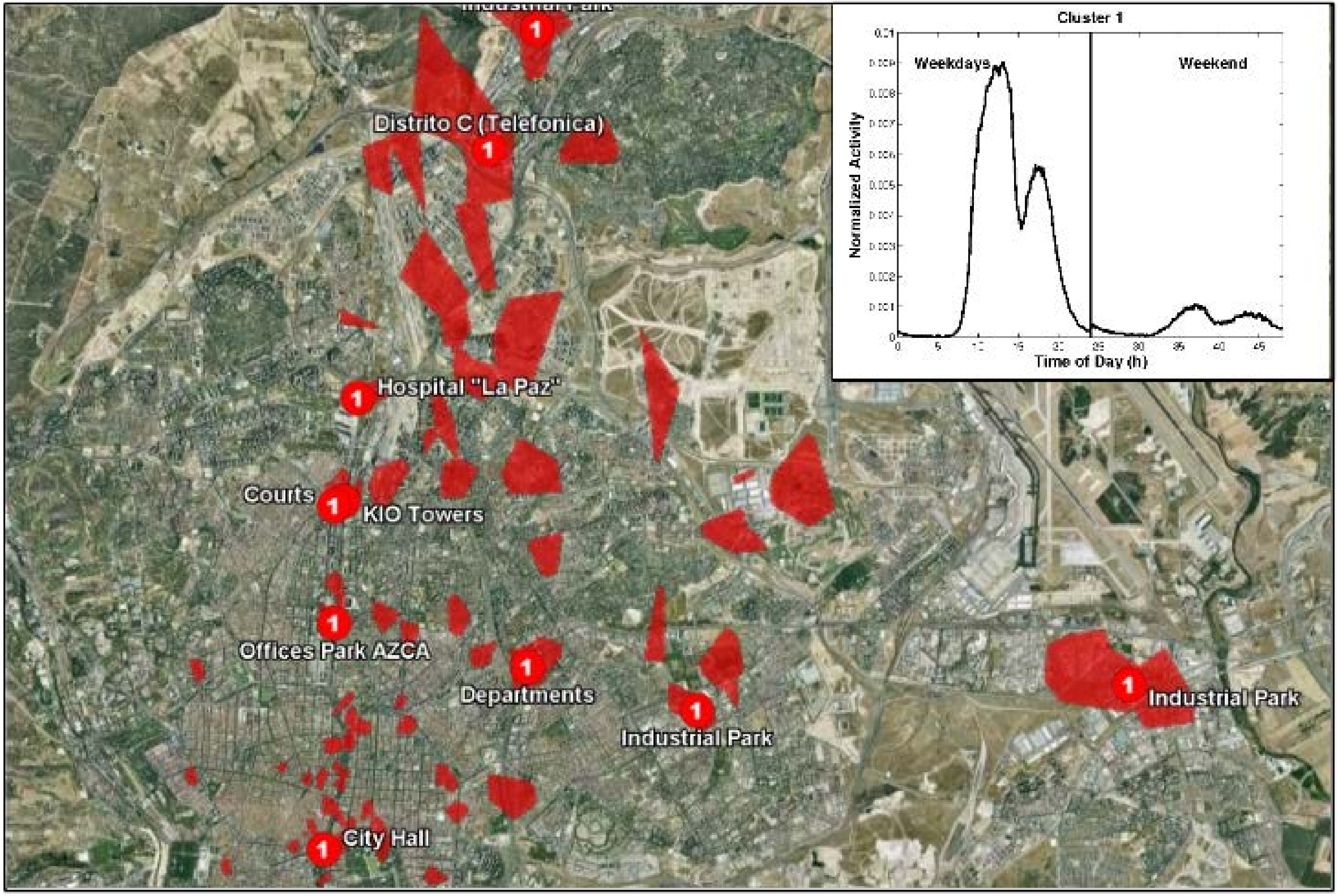
TARGET


CALIFICACIÓN
Residencial Multifamiliar
Residencial Unifamiliar
Residencial Mixto
Industria y Naves Agropecuarias
Terciario y Comercial
Equipamiento
Deportivo
Verde Público
Servicios
Infraestructuras
y Serv. Urbanos
Vías Pecuarias
M A D R I D
CIUDAD UNIVERSITARIA

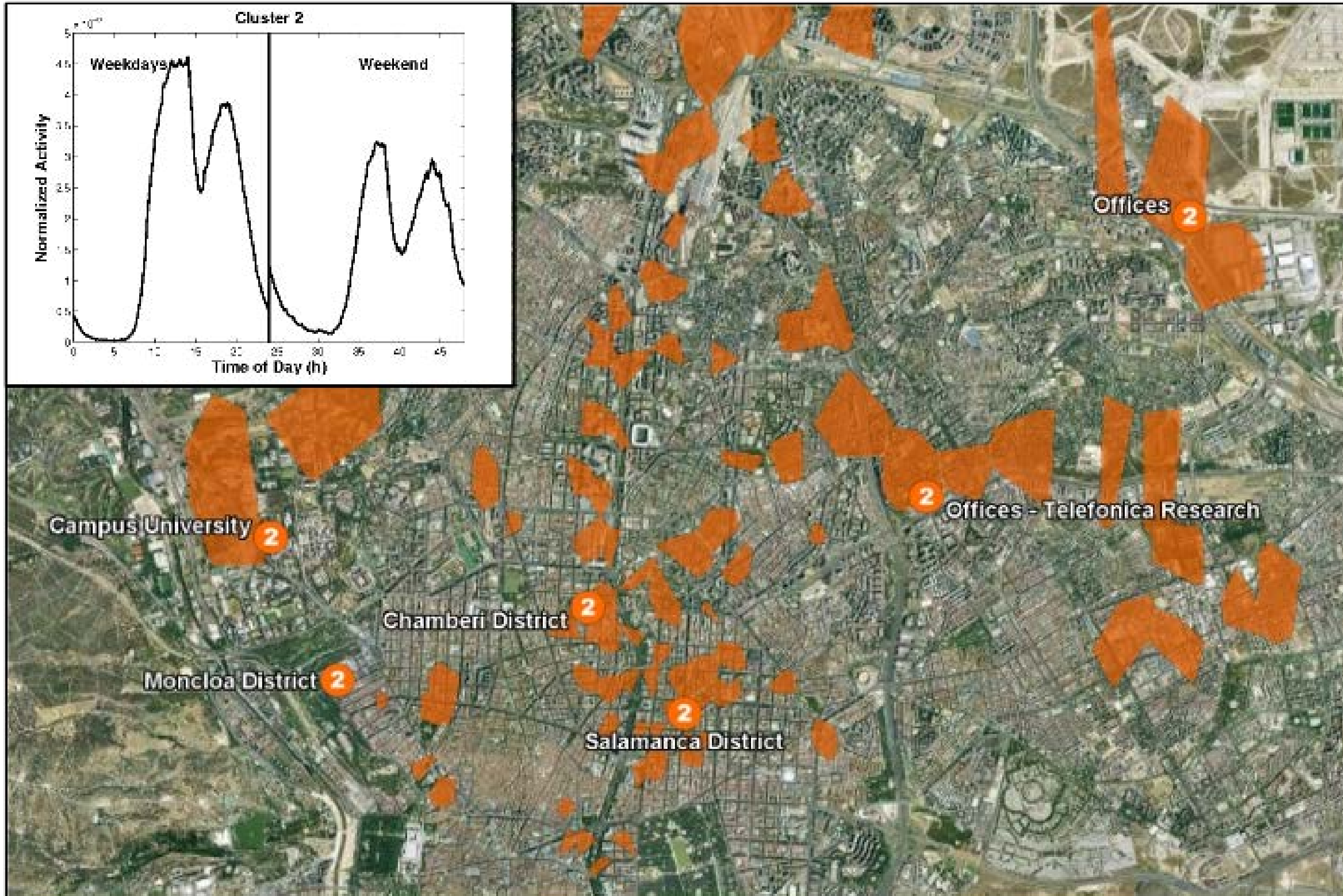


ID: 192519
ID: 192515
ID: 192516
ID: 215674
ID: 50664
ID: 266388
ID: 50670
ID: 266387
ID: 199917
ID: 269492
ID: 231488
ID: 199921
ID: 204489
ID: 215660
ID: 199923
ID: 199918
ID: 199919
ID: 340545
ID: 267963
ID: 267964
ID: 215662
ID: 206539
ID: 215636
ID: 50576
ID: 50657
ID: 97128
ID: 122693
ID: 215678
ID: 50662
ID: 50687
ID: 50665
ID: 483651
ID: 266284
ID: 266283
ID: 269205
ID: 189533

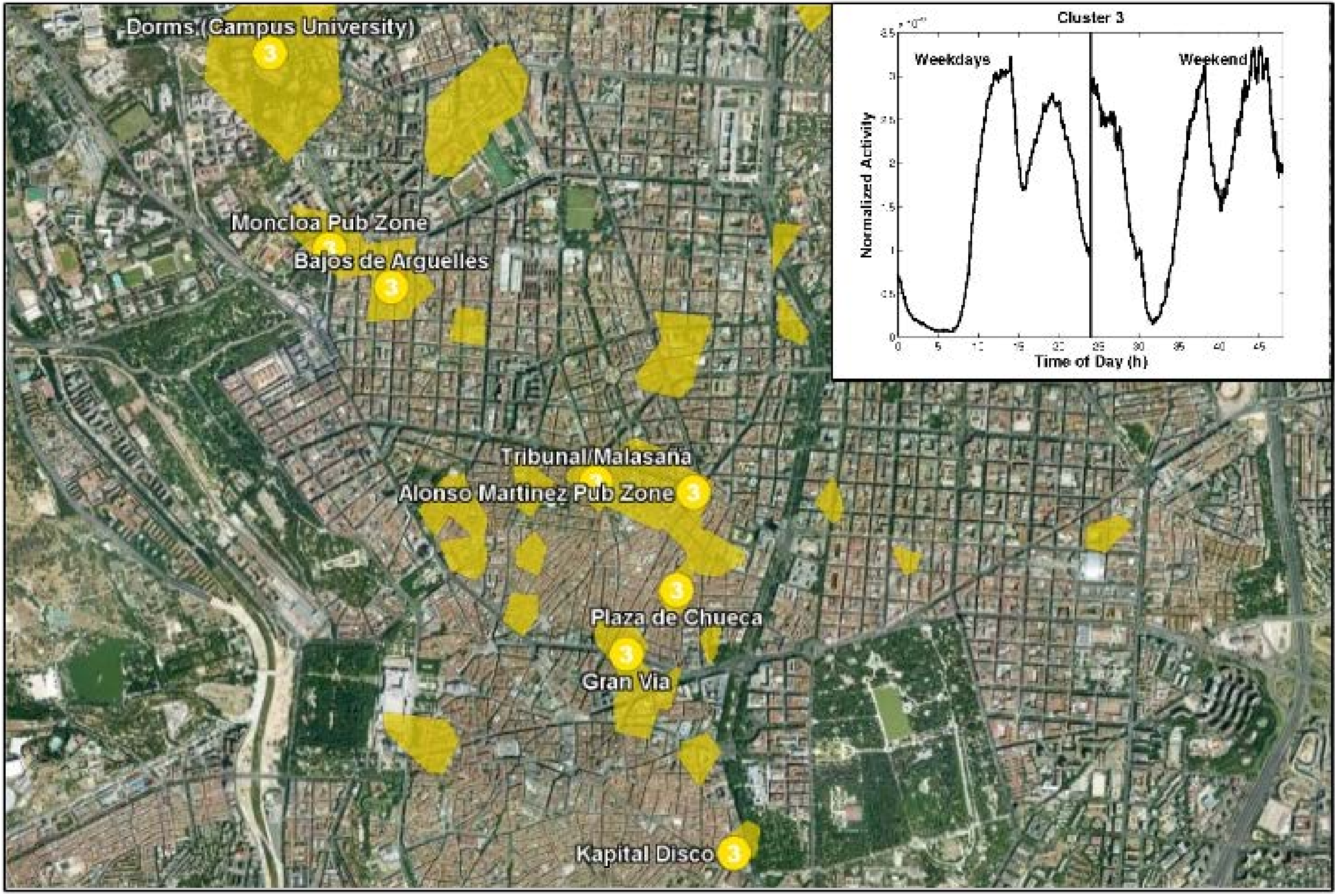
Cluster 1: Industrial & Office



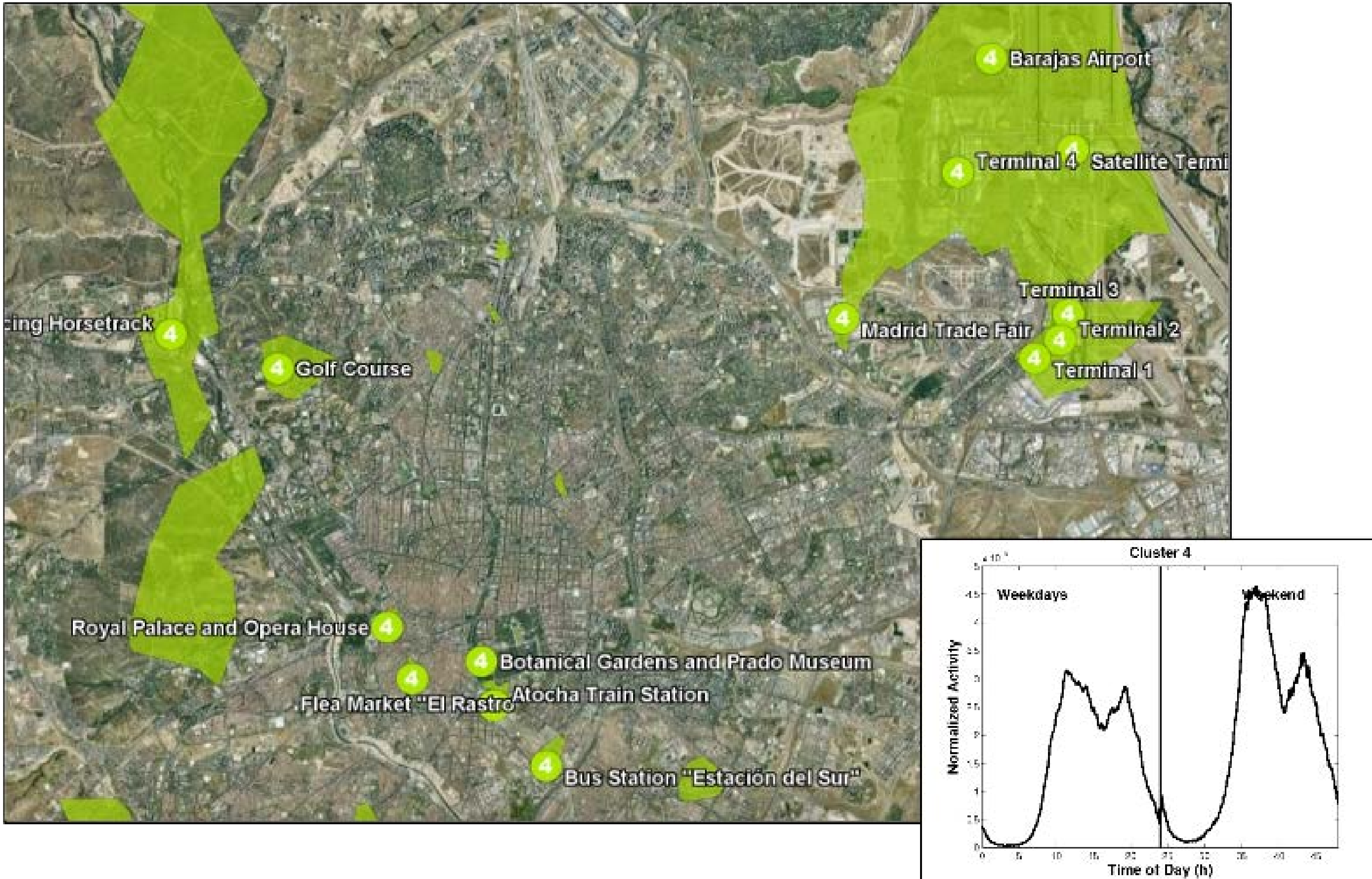
Cluster 2: Business & Commercial



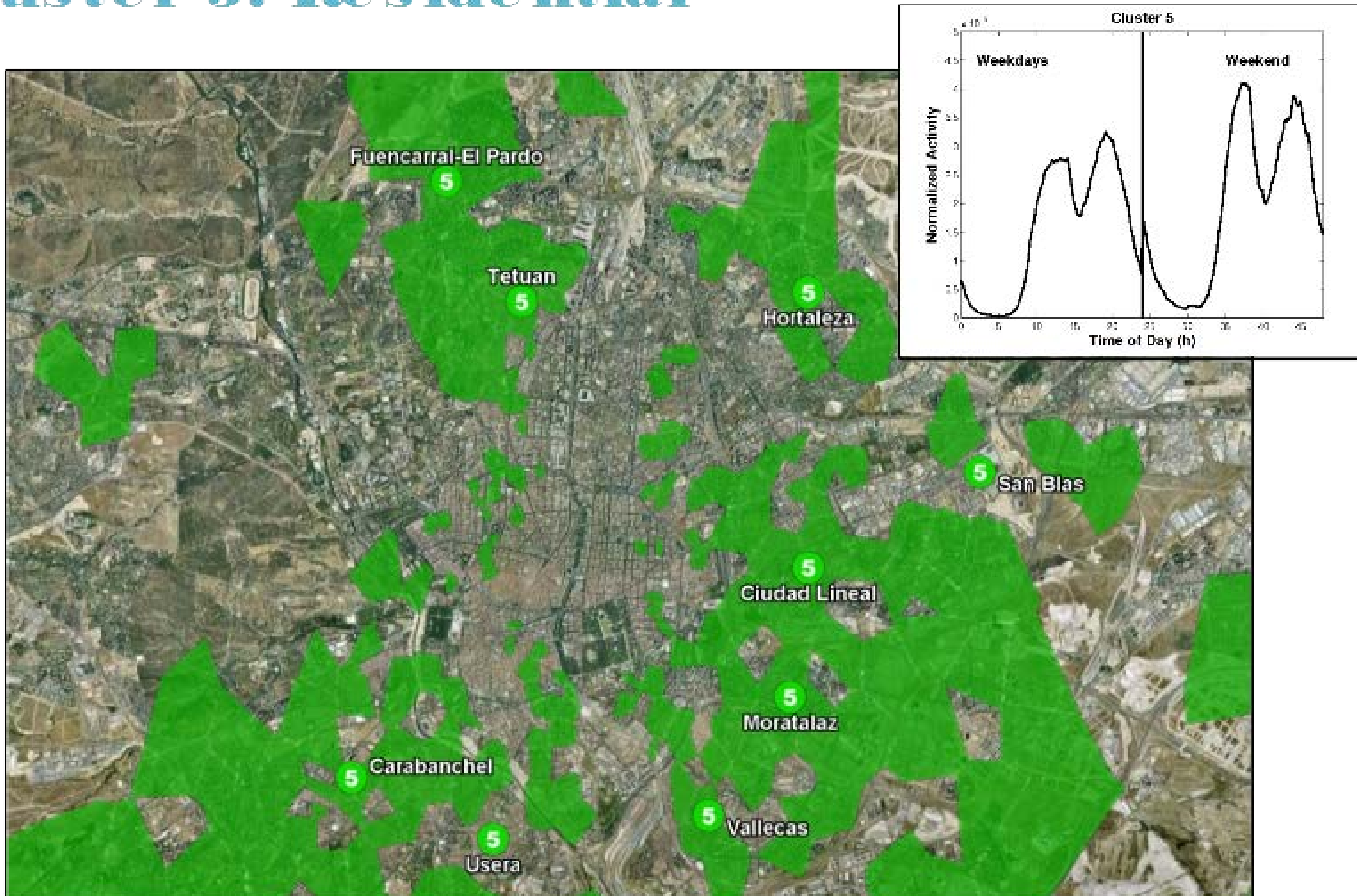
Cluster 3: Nightlife

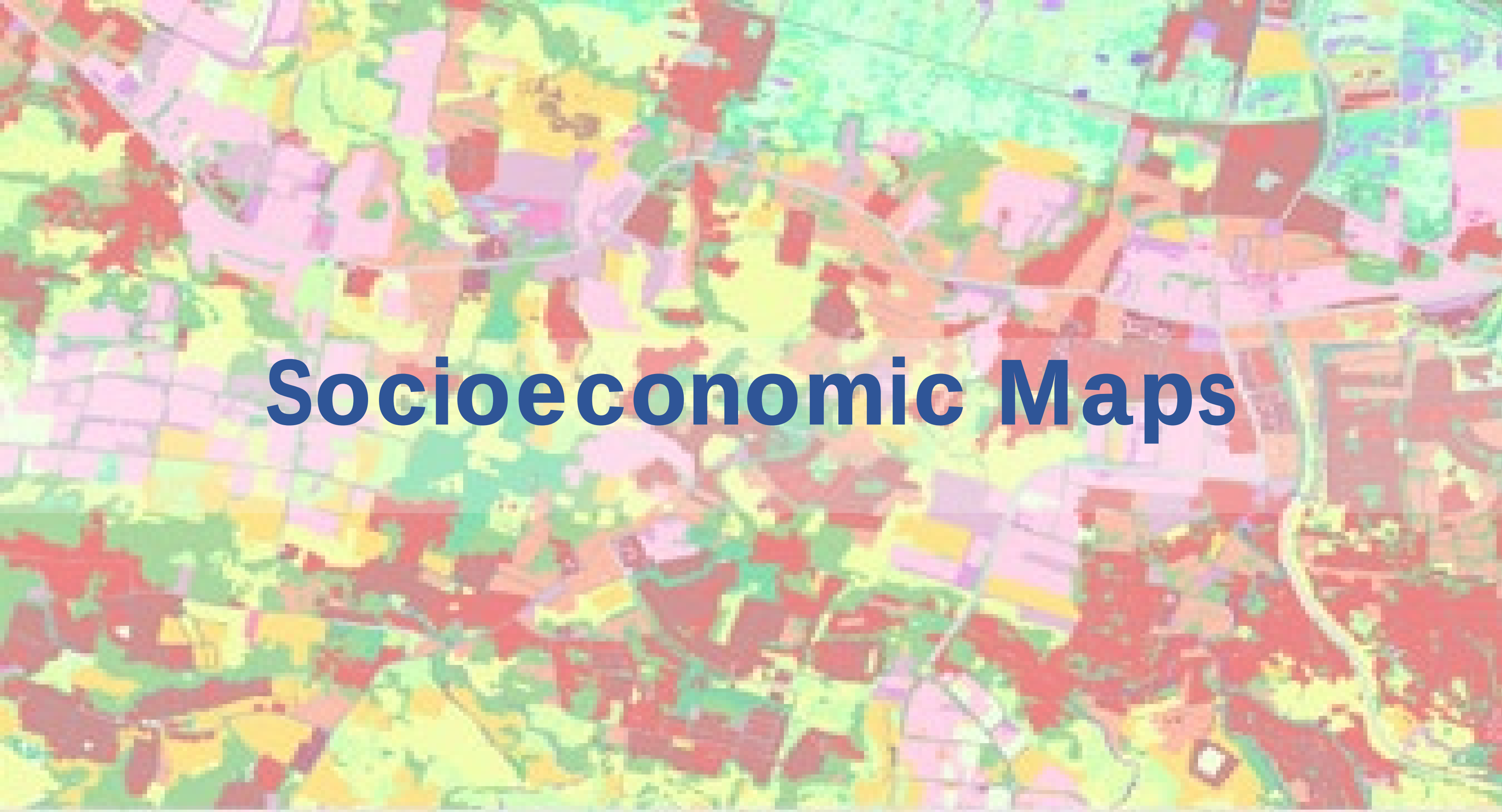


Cluster 4: Leisure & Transport



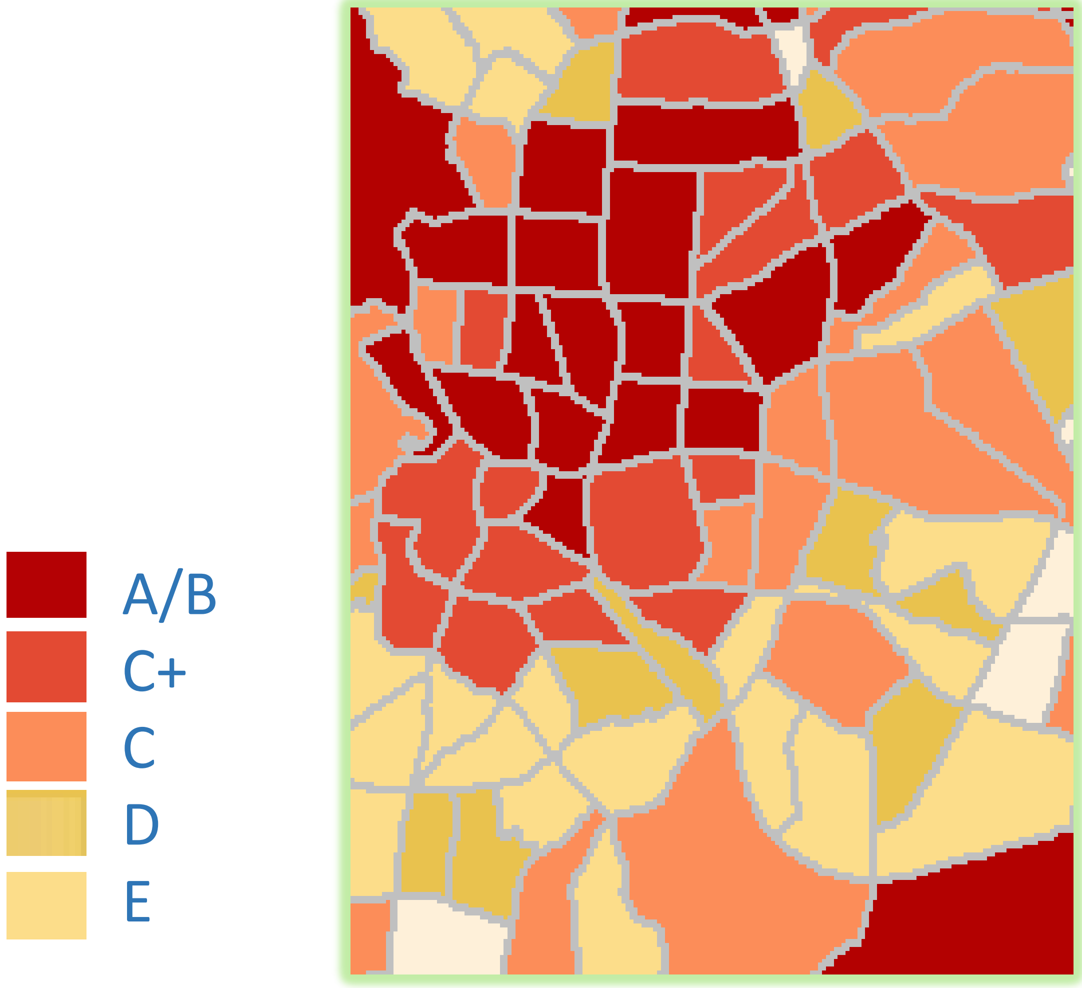
Cluster 5: Residential



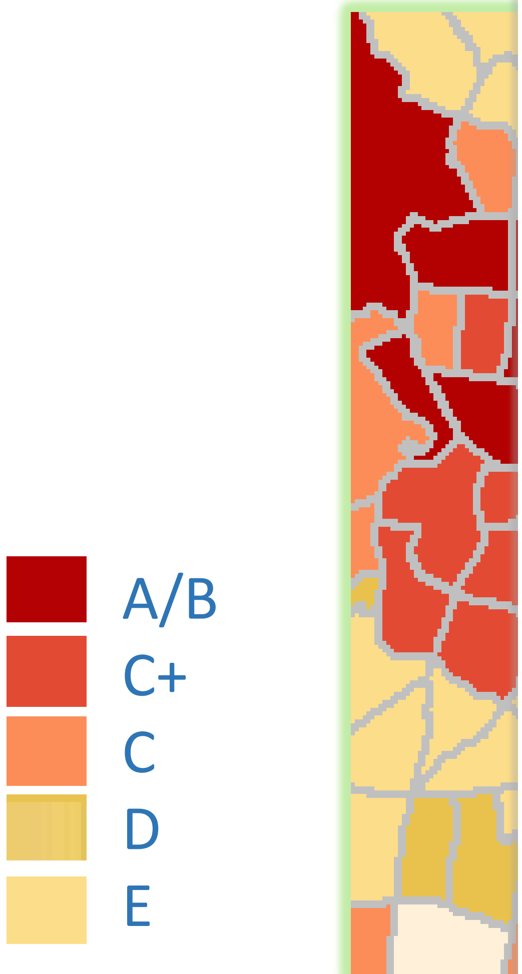


Computing Socioeconomic Maps from Cell Phone Data

Motivation: Socioeconomic Maps

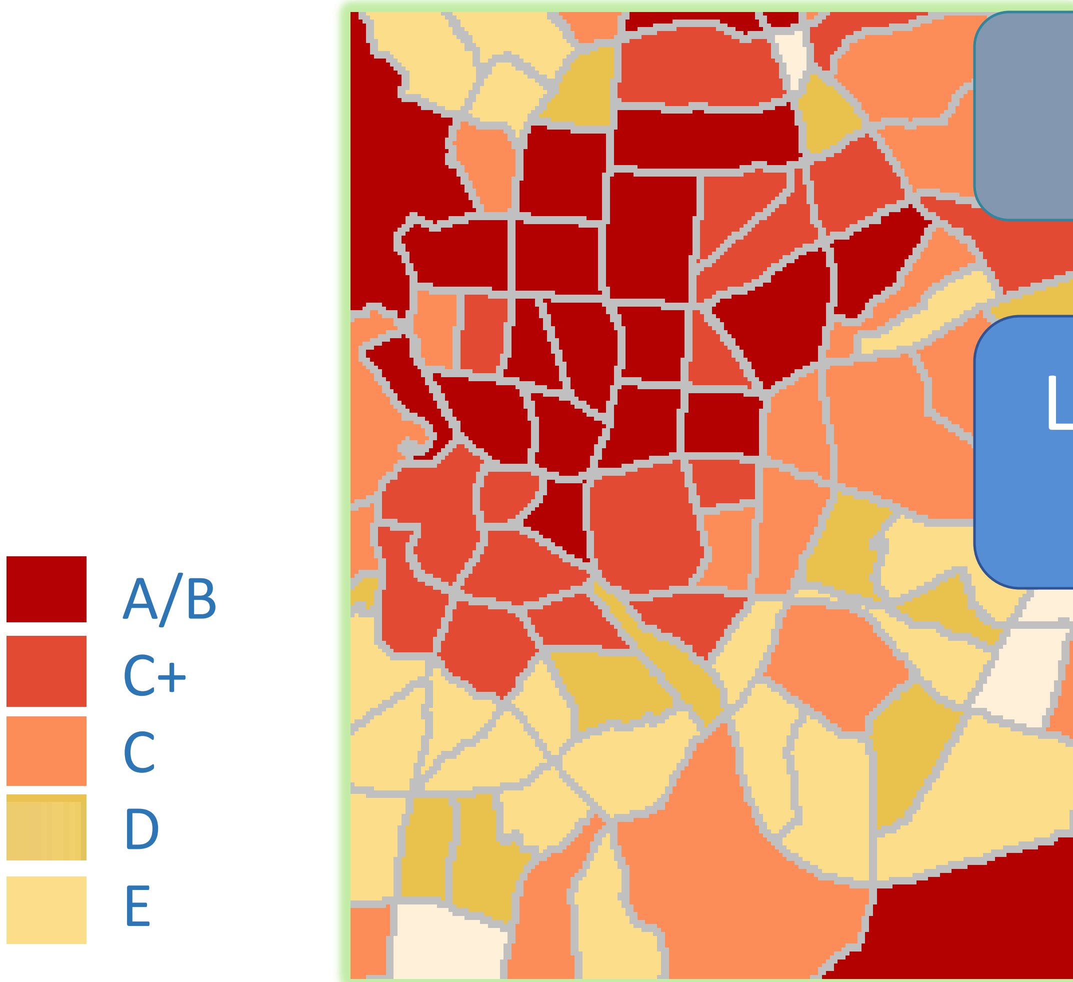


National Statistical Institutes (NSI)



Census Variables	
Variable Type	Description
Education	% of Population with Primary School
	% of Female Population with Primary School
	% of Male Population with Primary School
	% of Population with Secondary School
	% of Female Population with Secondary School
	% of Male Population with Secondary School
	% of Illiterate Population
	% of Female Illiterate Population
	% of Male Illiterate Population
Demographics	% of Female Population
	% of Male Population
	% of Young Population (< 16)
	% of Middle-Age Population ($16 - 60$)
	% of Senior Population (> 60)
Goods	% of Houses with Cement Floor
	% of Houses with 1 room
	% of Houses with 3+ rooms
	% of Houses with Electricity
	% of Houses with Water
	% of Houses with TV
	% of Houses with PC
	% of Houses with All
SEL	Socio-Economic Level

Important Data Comes at a Price



Can human behaviors extracted from Call Detail Records be used to forecast regional socioeconomic information?

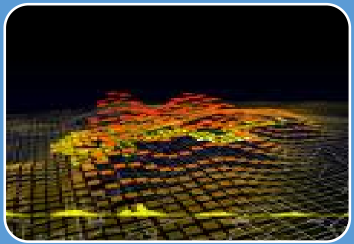
Modeling Human Behavior

Consumption

- Number calls, duration, frequency, SMS/MMS/voice
- Expenses
- Handset Type and Features

Social

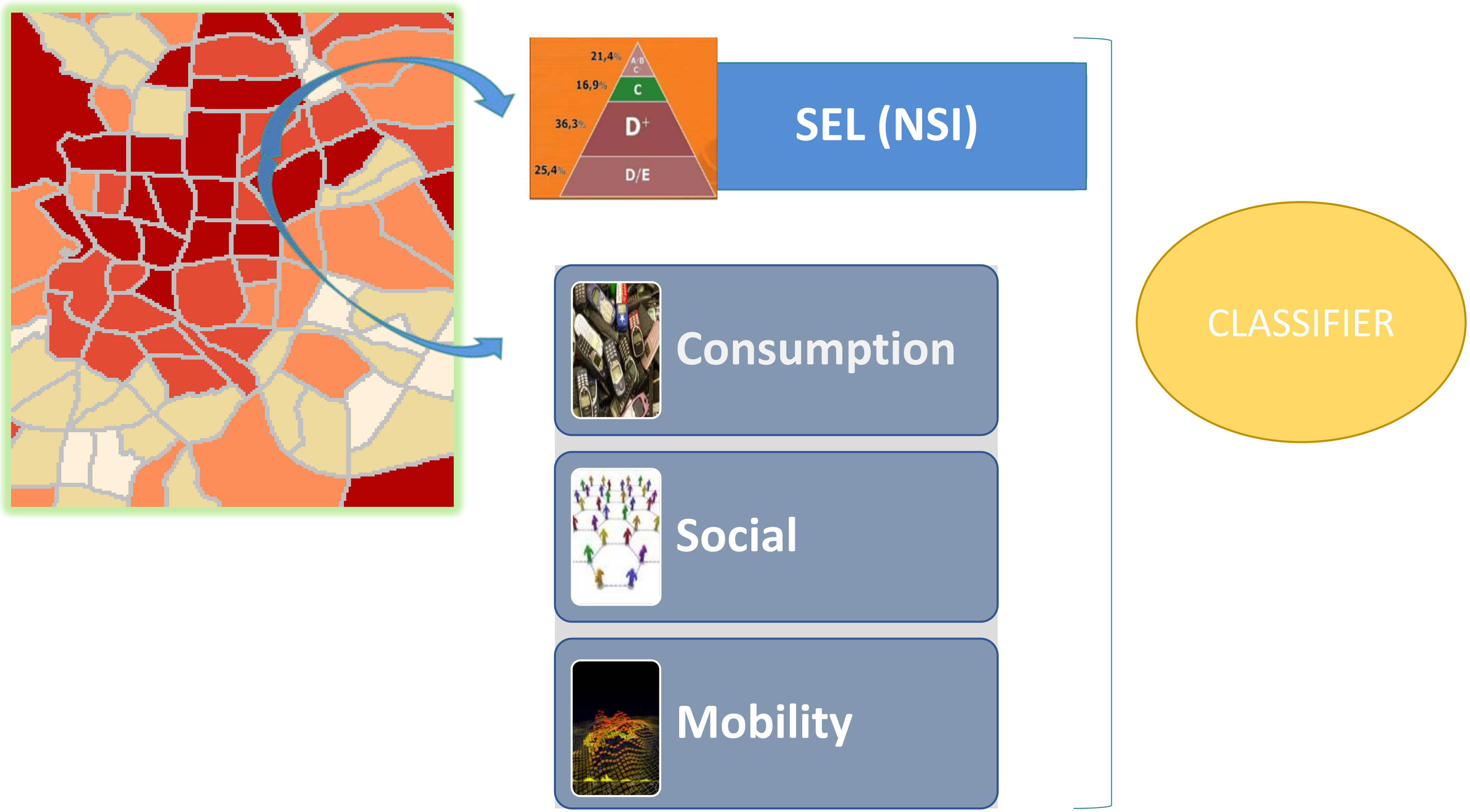
- Degree of the social network
- Strength of the contacts (Reciprocity & Frequency)
- Geography of the social contacts



Mobility

- Mobility Patterns (Entropy)
- Diameter of mobility
- Radius of gyration (Home/Work)

Methodology



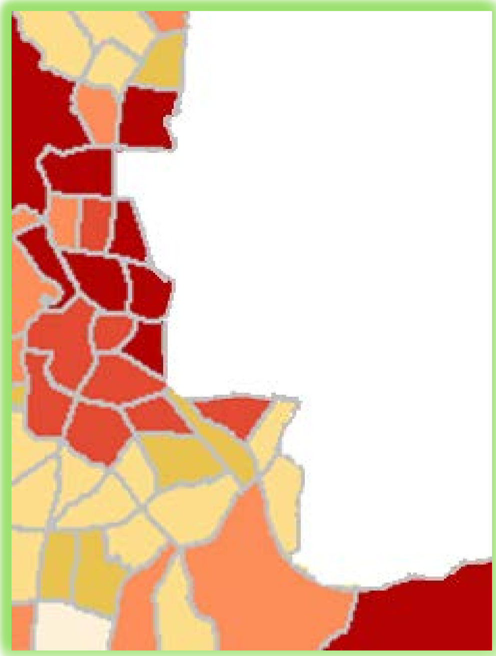
Cost-effective Maps

NSI surveys subset of regions

Call Detail Records

Forecasting Models

Predict the Present



REDUCE COSTS

Datasets

- Data for a city in Latin America (NSI)
  - 1200 regions (GUs)
  - SEL values from 0..100
- Call Detail Records
  - 6 months, 500K customers
  - City has 920 coverage areas
  - 279 variables per coverage area

Evaluation Results (6 SELs)

EM Clustering
P=62%
20 features , 6 classes

Random Forests
P=60%
22 features

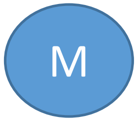
SVM
P=57%
19 features

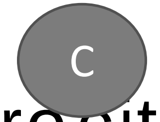
CLASSIFICATION

	A	B	C	D	E	F
A	**0.59**	0.41	0.00	0.00	0.00	0.00
B	0.10	**0.60**	0.30	0.00	0.00	0.00
C	0.00	0.20	**0.66**	0.14	0.00	0.00
D	0.00	0.00	0.21	**0.64**	0.08	0.07
E	0.00	0.00	0.00	0.20	**0.60**	0.20
F	0.00	0.00	0.00	0.13	0.35	**0.52**

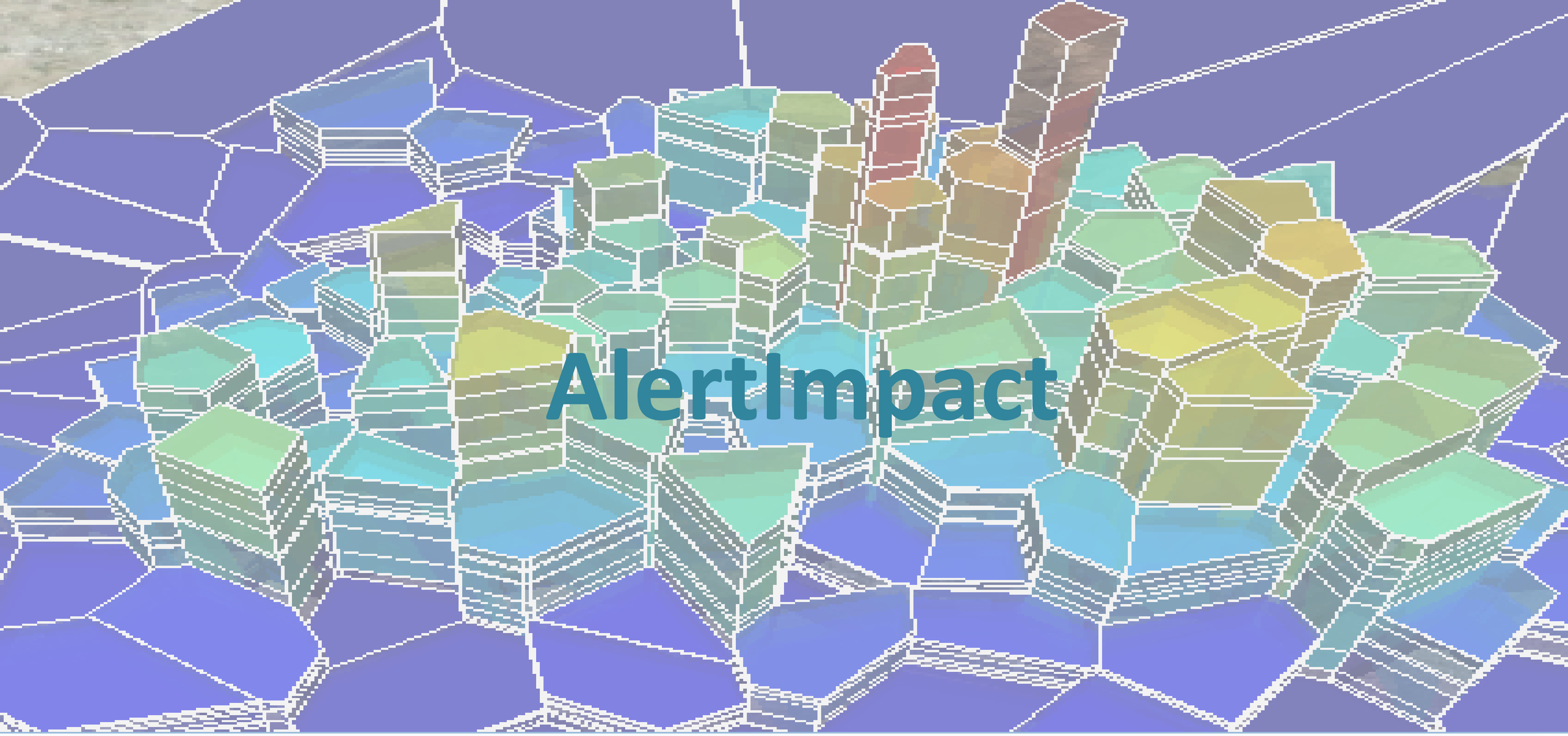
Results superior to baseline
No spatial autocorrelation

Most predictive features…

- Weekly average of different BTSs used


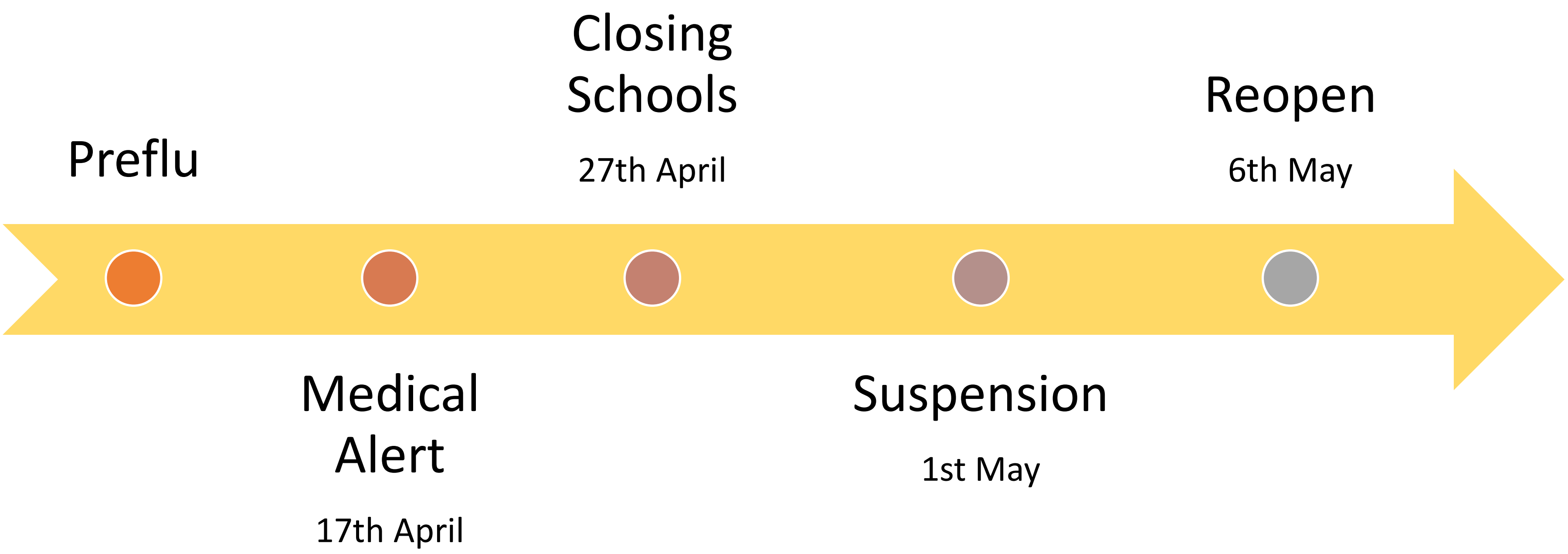
- Percentage of incoming SMS over all communications


- Percentage of SMS sent with high reciprocity

Understanding the Impact of Health Alerts using Cell Phone Data

H1N1 Mexico Timeline



Can we measure the effect that government alerts had on the mobility of the population using Call Detail Records?

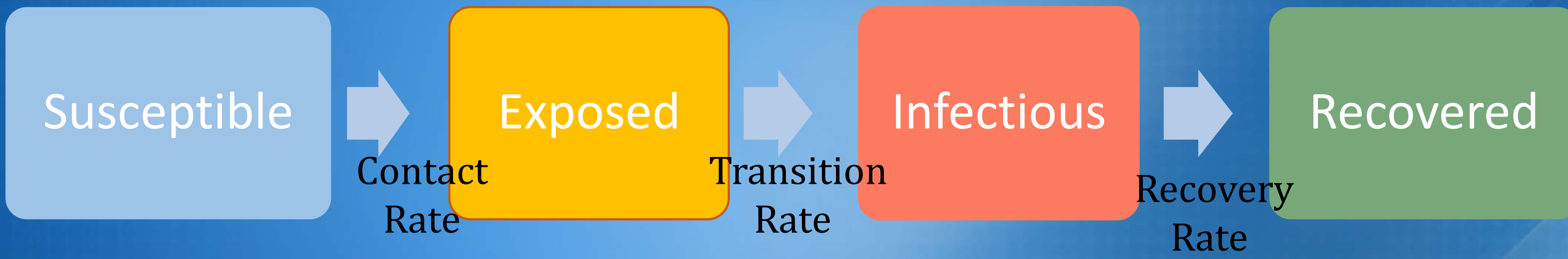
(citizens do what they are asked to do)

Can we measure the effect that government alerts had on the spreading of the epidemic using Call Detail Records?

(government measures are useful)

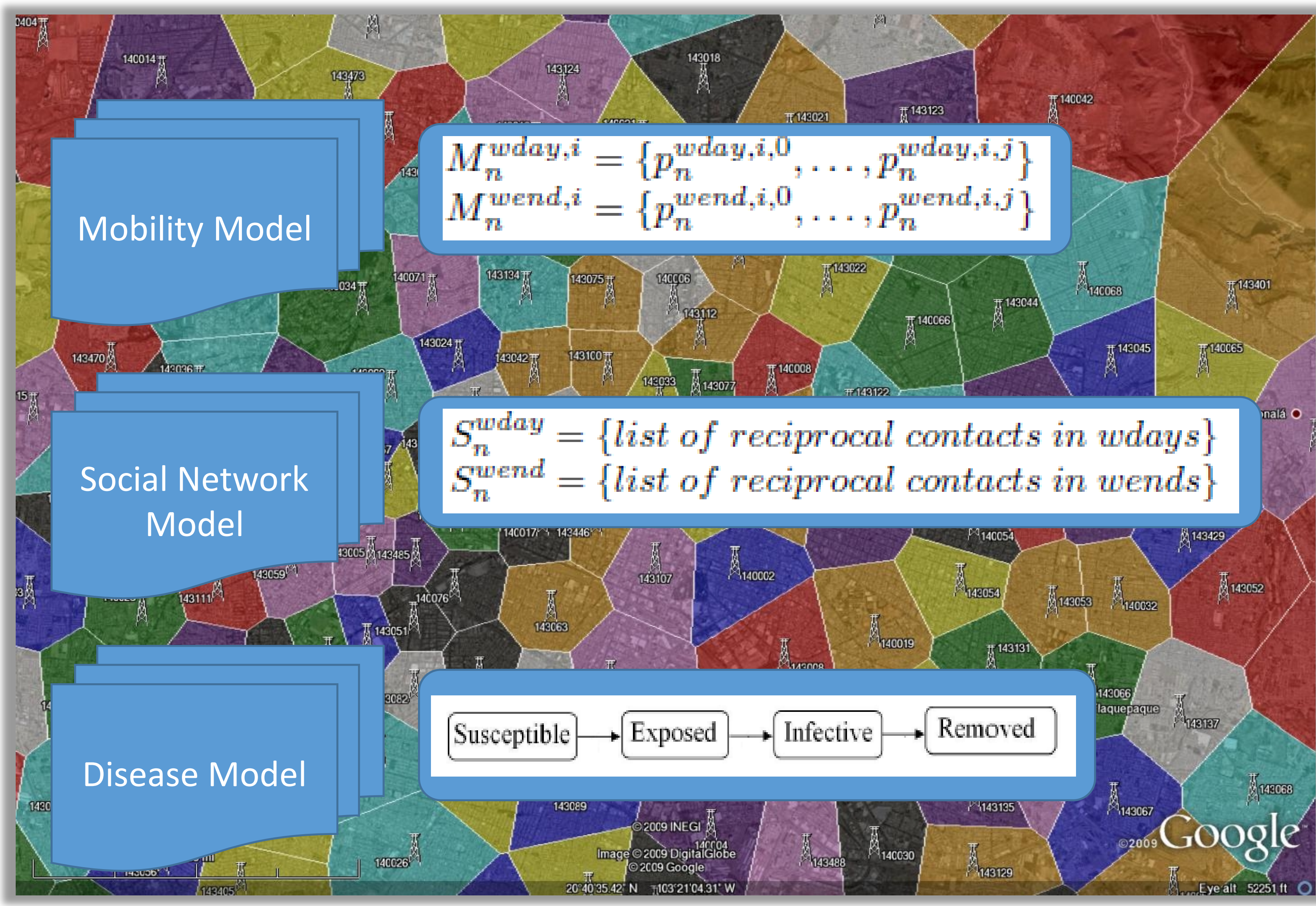
Epidemic Disease Models

- Compartmental Models (SEIR)

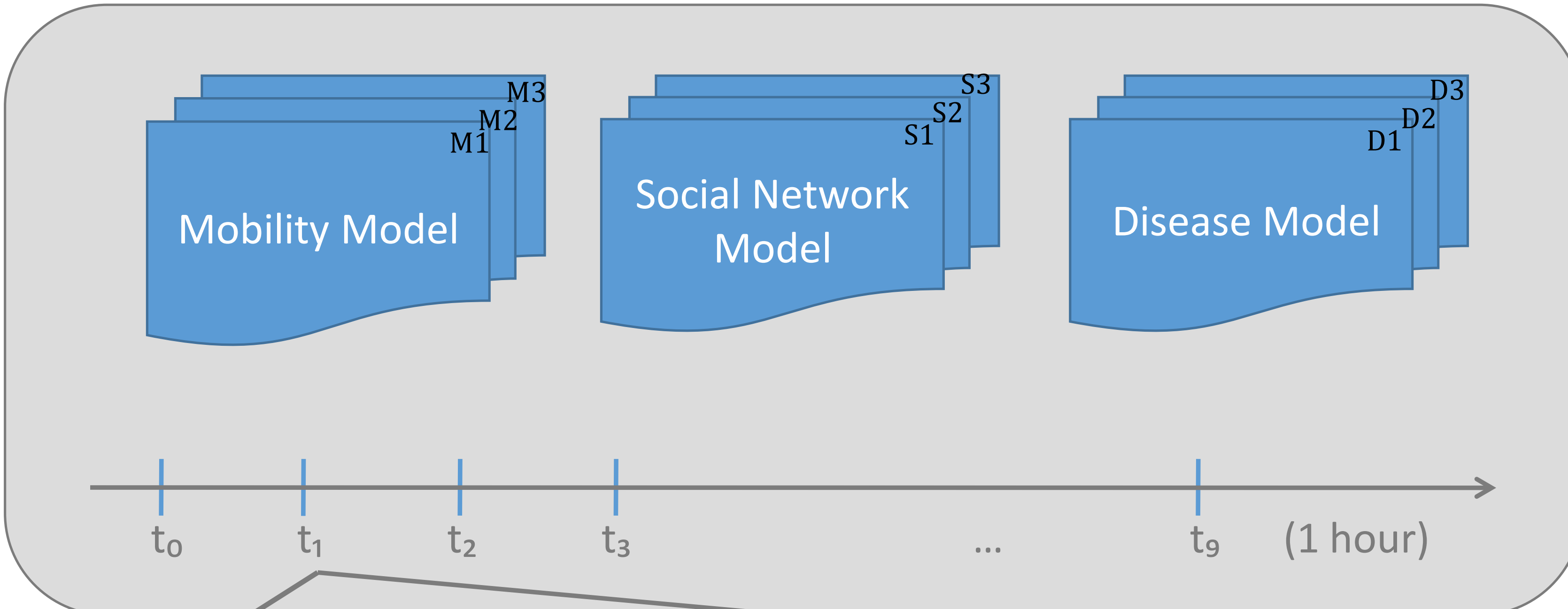


- Agent Based Models
  - Capure complexity of social interaction
  - Limitation with the information available to generate the agents

ABM for Virus Spreading using CDR

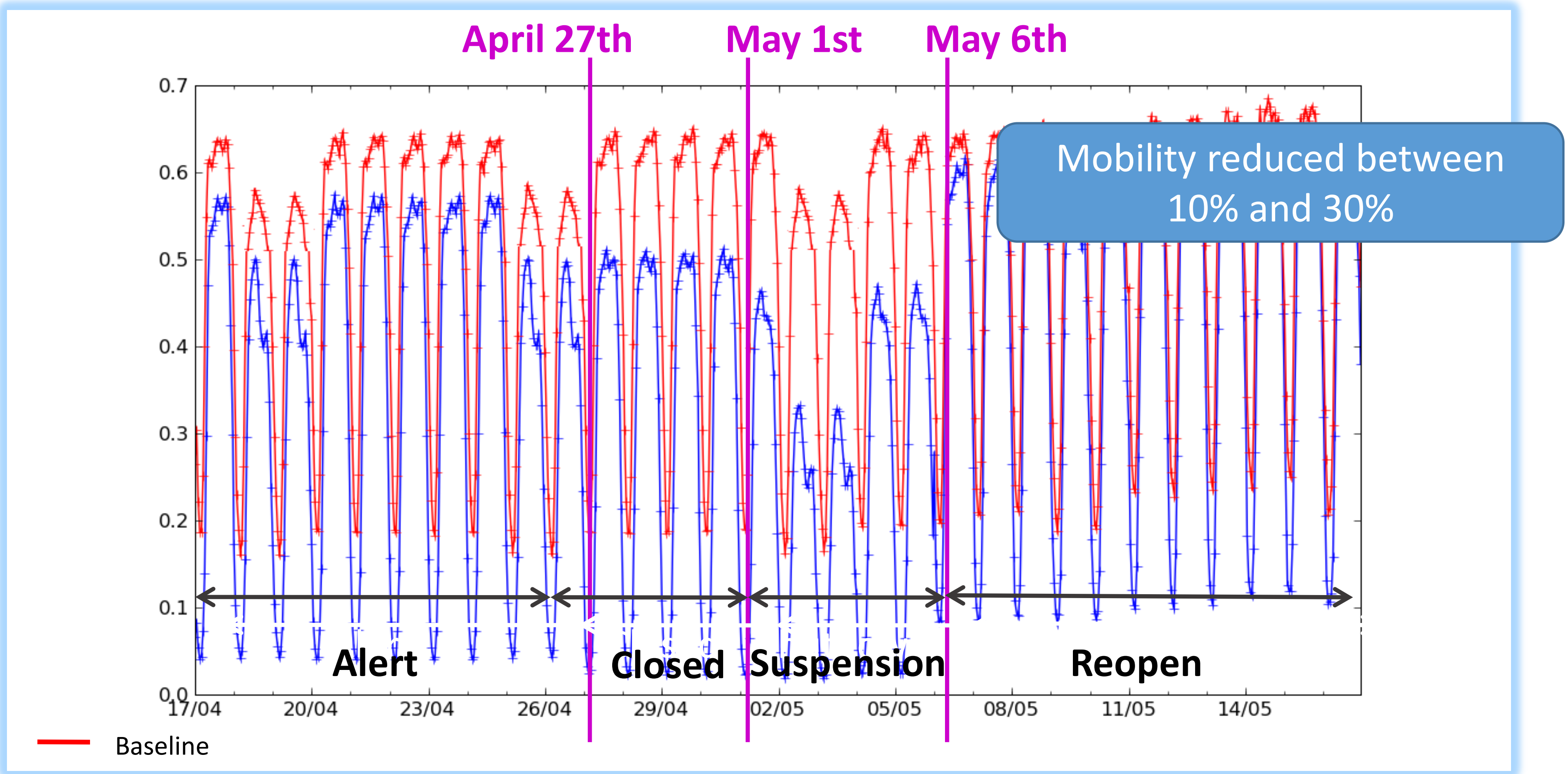


Discrete Event Simulator

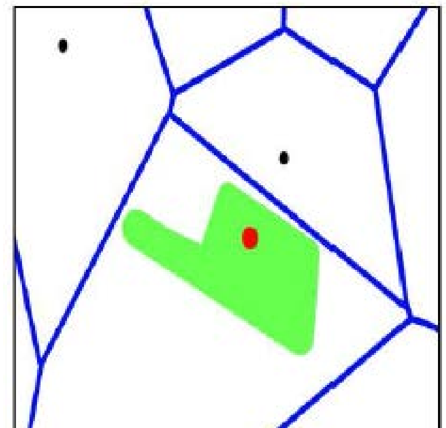


Identify geographical location (BTS)
Identify peers in same BTS
  If peer in SN then evolve disease model with p_i
  Else evolve disease model with p_j

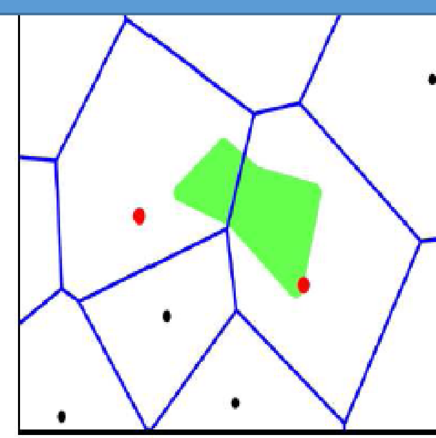
Changes in Mobility



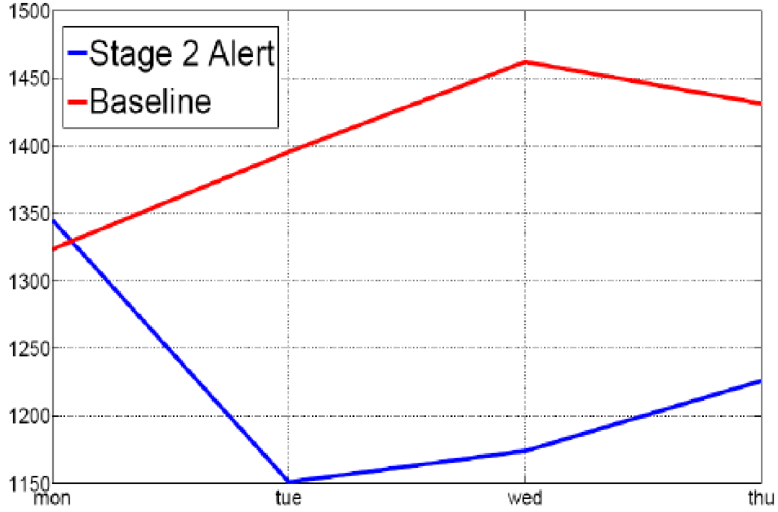
Infrastructure Analysis

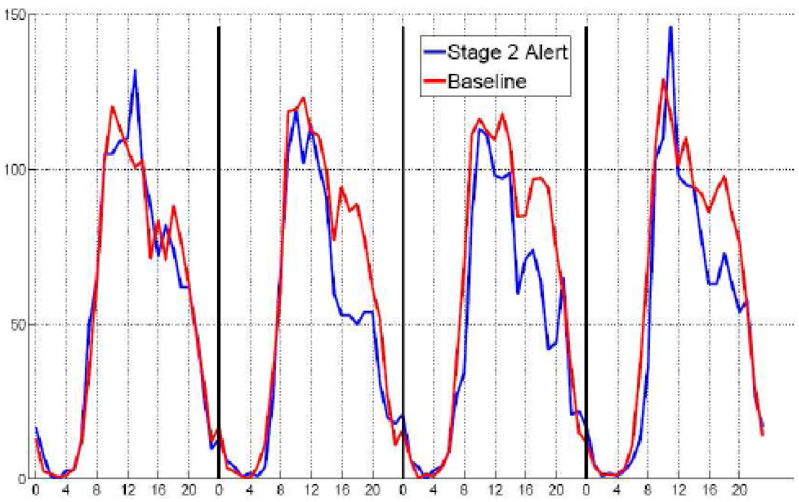


Only citizens that DO NOT LIVE in these BTS areas



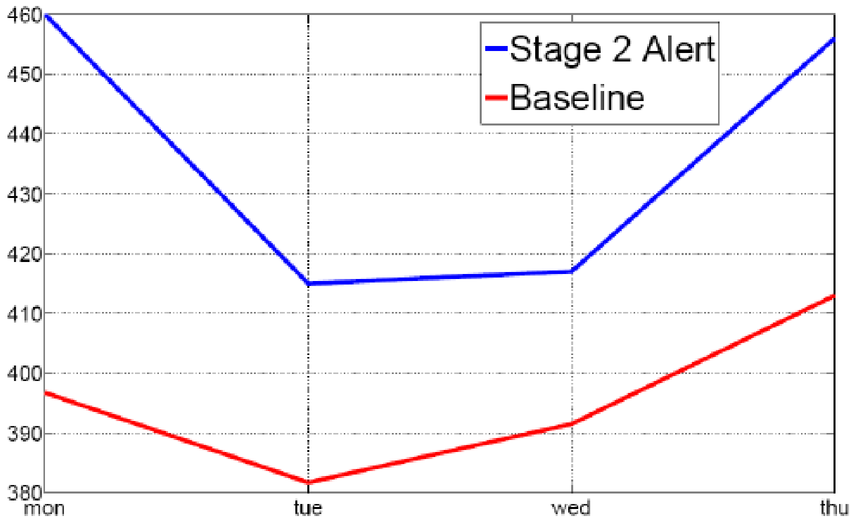
University Campus

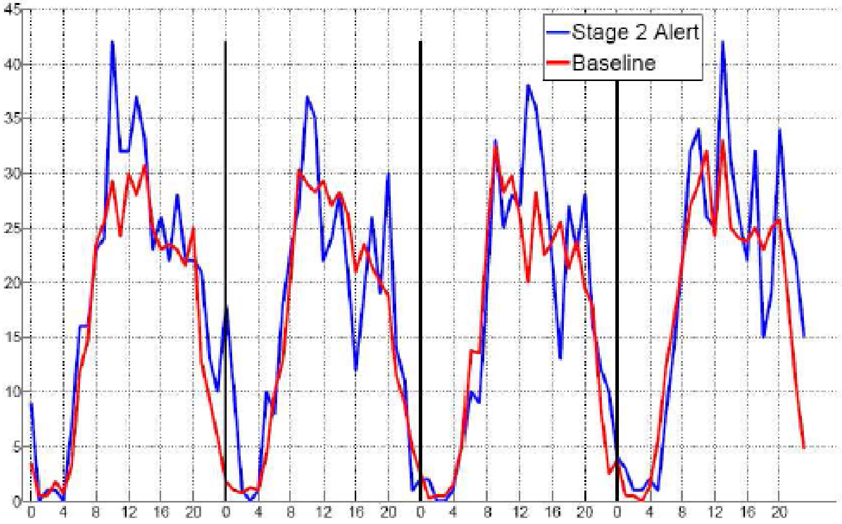




Statistically Significant Decrease during Stages 2 and 3

Airport

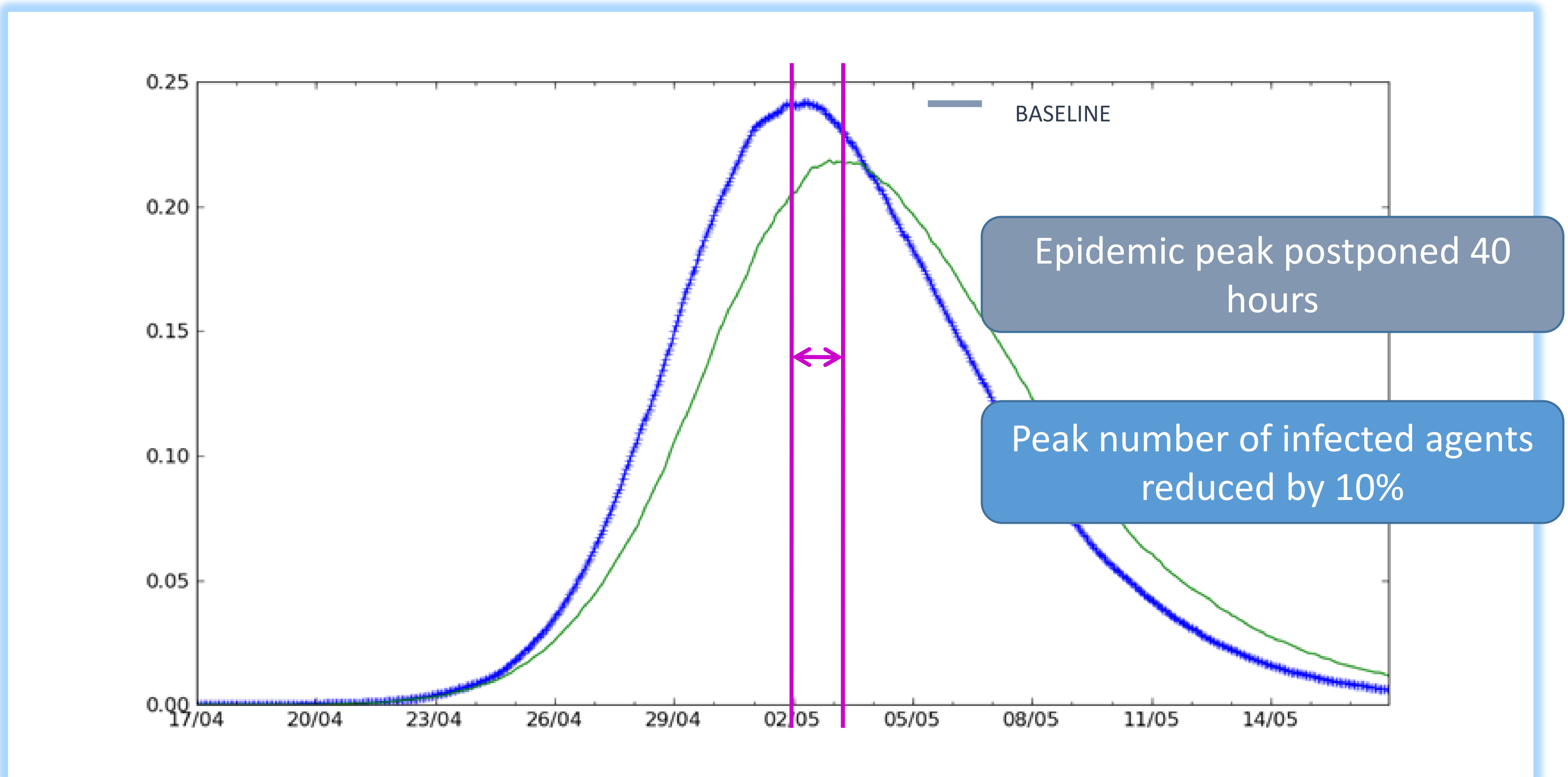




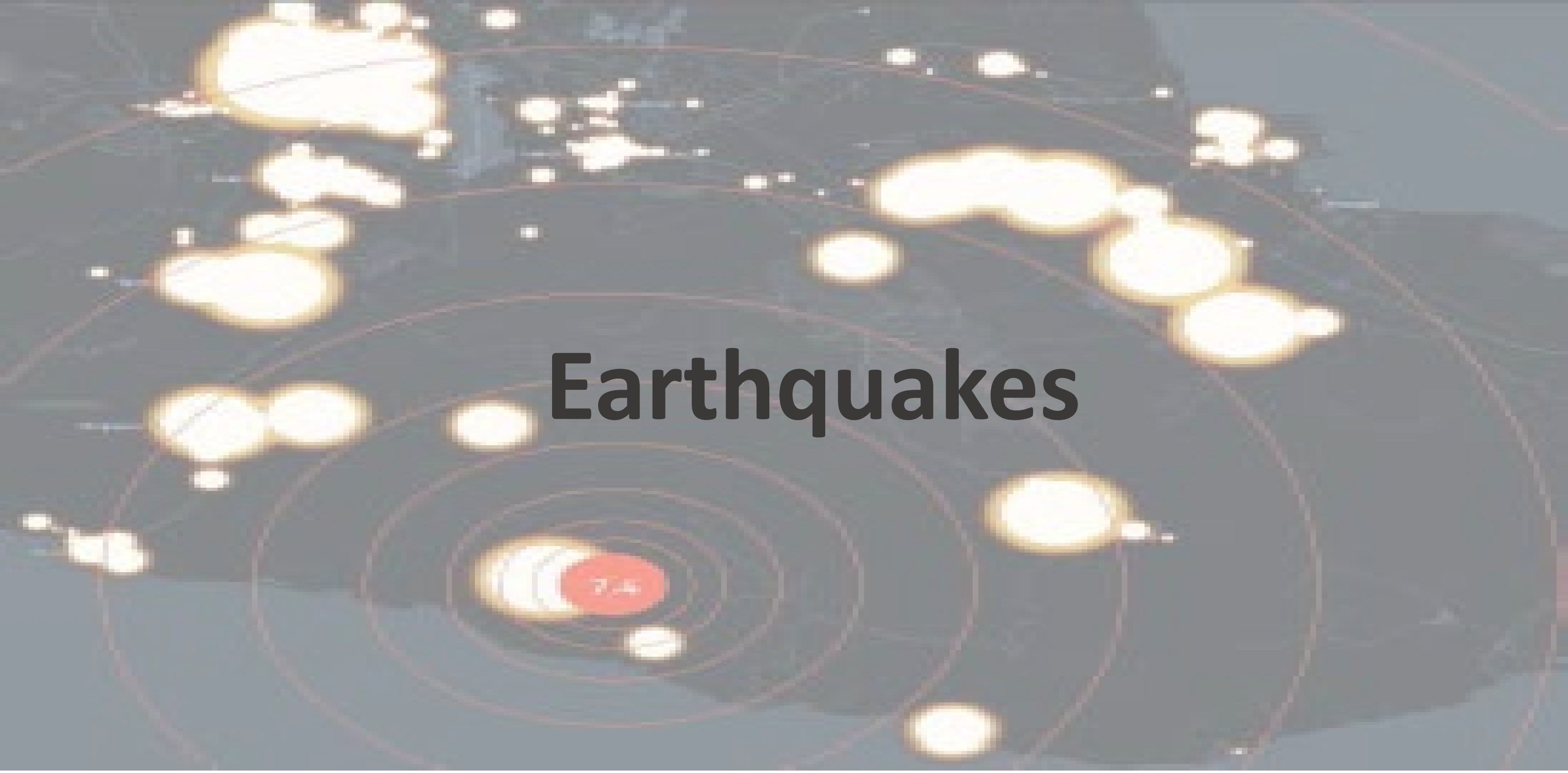
Statistically Significant Increase during Stages 2 and 3

Effect on Epidemic Spreading

Evaluation

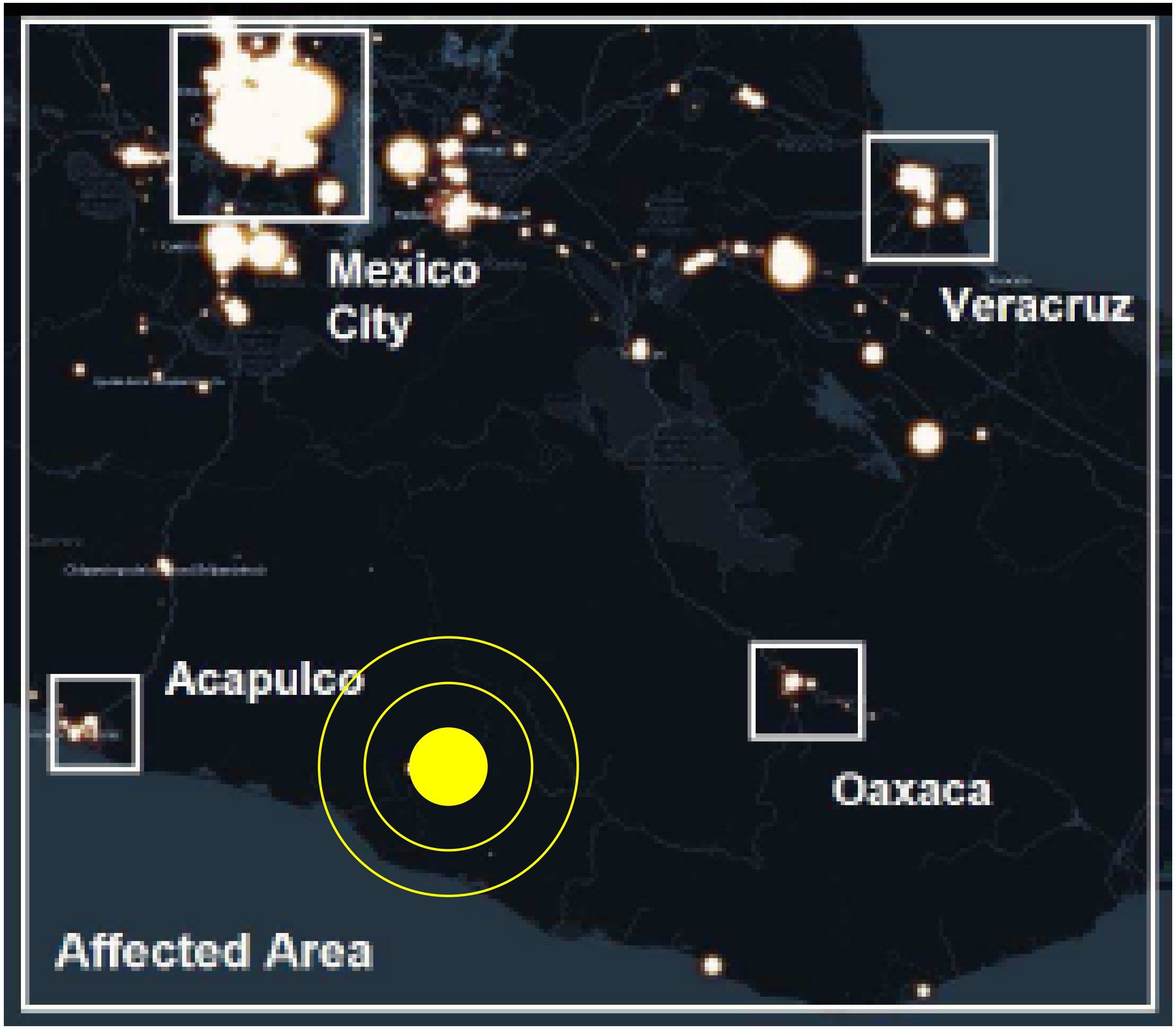

0.25
0.20
0.15
0.10
0.05
0.00
17/04
20/04
23/04
26/04
29/04
02/05
05/05
08/05
11/05
14/05
BASELINE
Epidemic peak postponed 40 hours
Peak number of infected agents reduced by 10%

Earthquakes

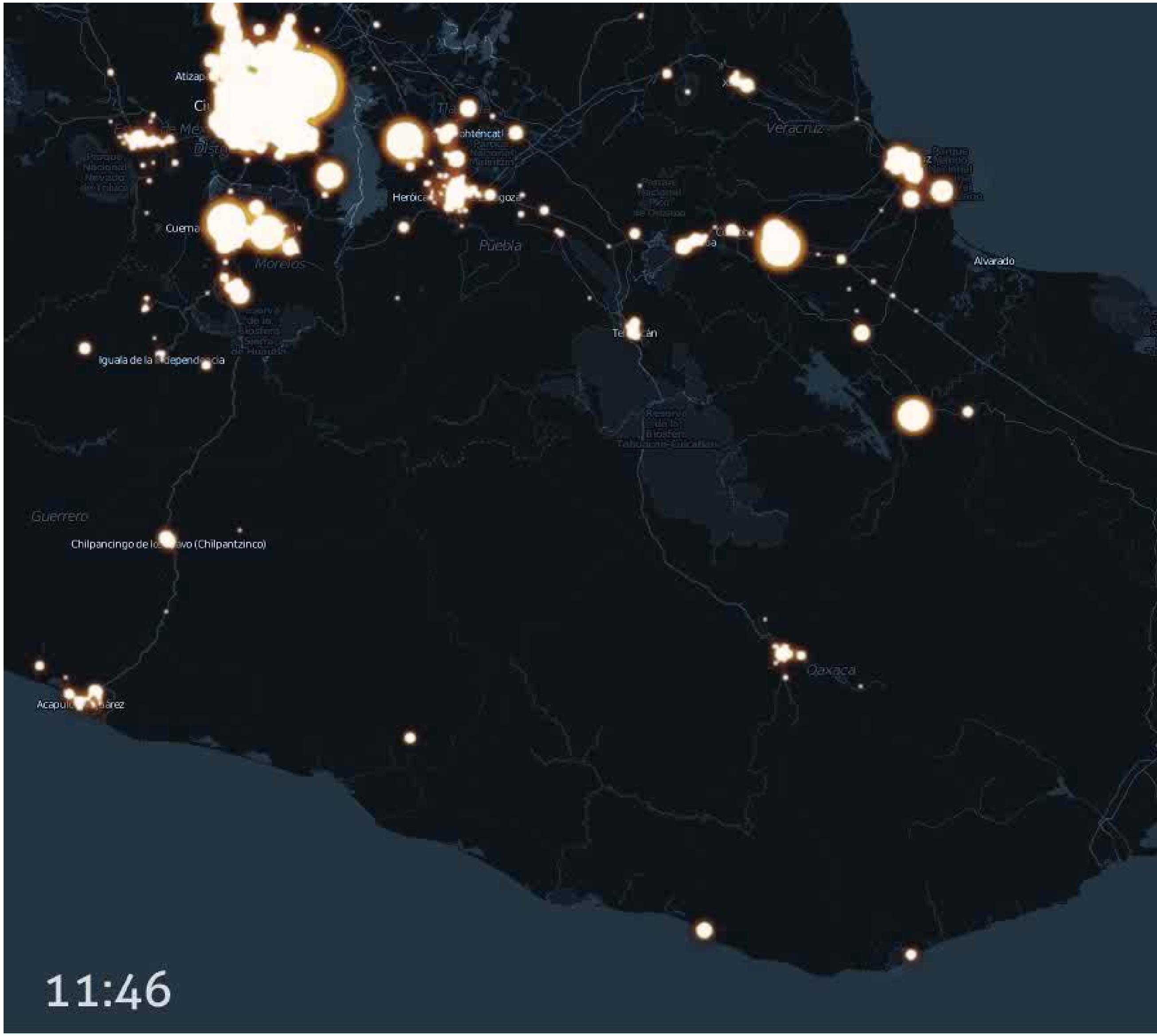


Characterizing the Social Response to an Earthquake

Oaxaca Earthquake



Can we measure the social response to the earthquake using call detail records?



Mexico 2012 Earthquake

Magnitude: 7.4

Date & Time: March 20, 2012 at 12:02:48 PM

Location: 16.662°N, 98.188°W

Depth: 20 km (12.4 miles)

Damage: 2 deaths

13 injuries

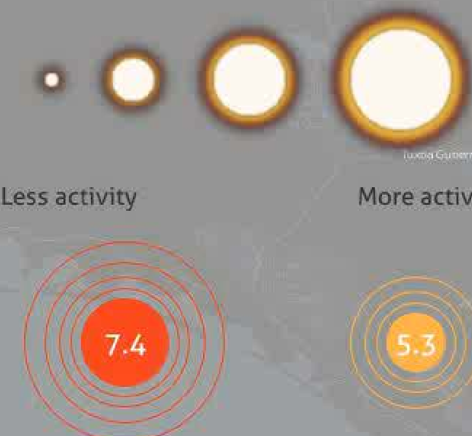
800 houses collapsed

Data: U.S. Geological Survey

Telefonica Mexico

OpenStreetMap

Legend



Less activity More activity

Earthquake Aftershocks

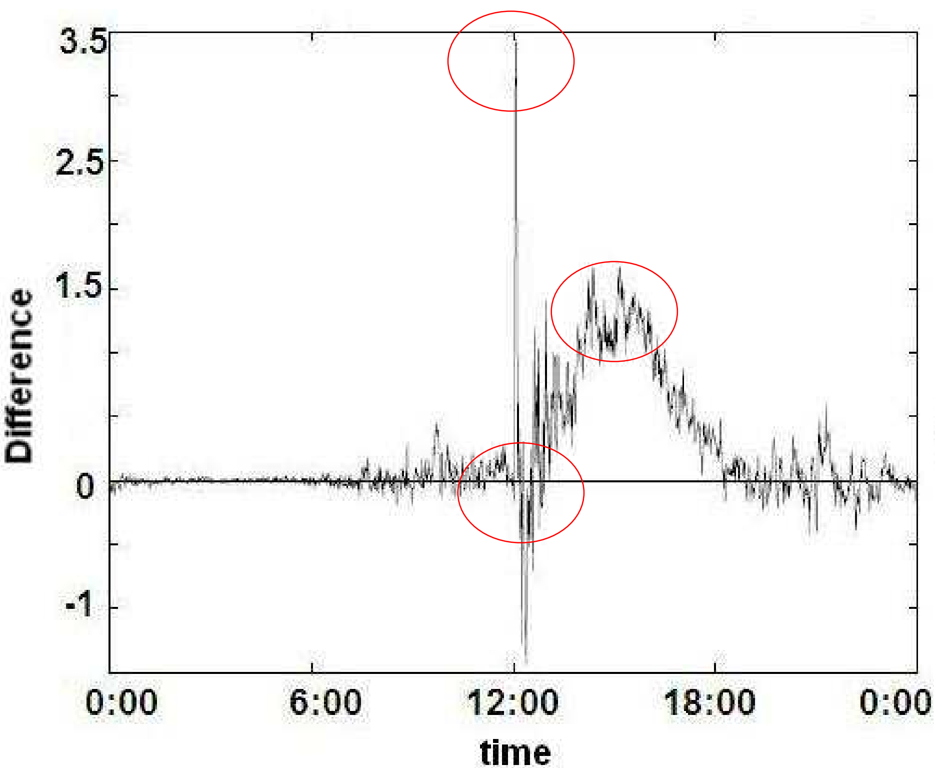
Methodology

- Focus on three social responses:
  - Communication: call volumes and call durations
  - Social activity
  - Mobility

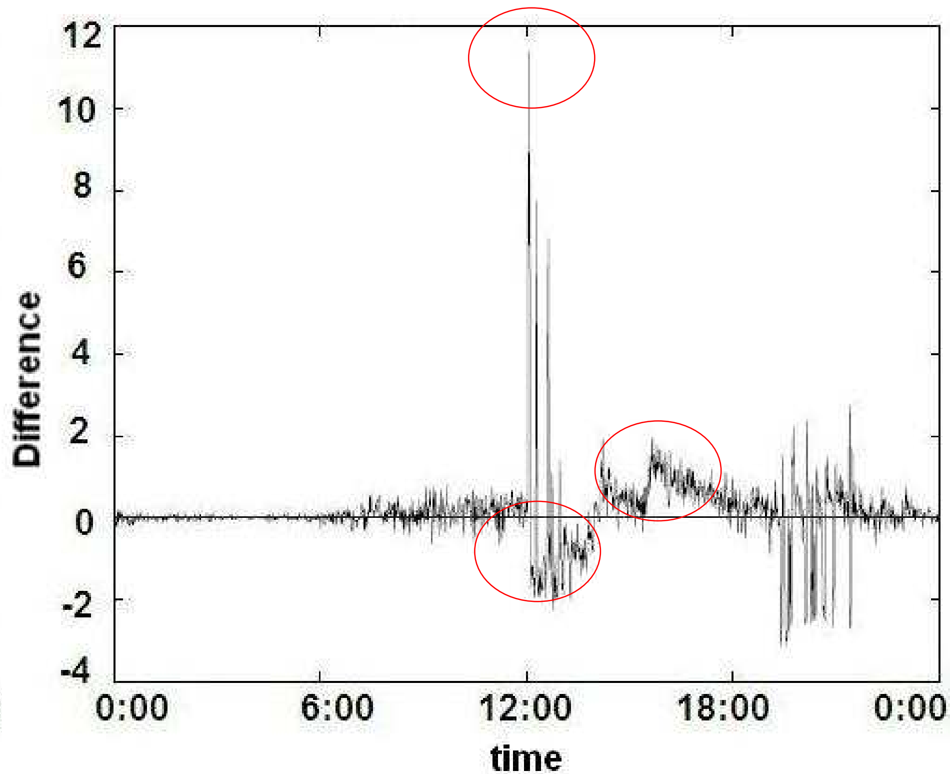
- Compare responses across cities
  - Day of the earthquake
  - Baseline

Call Volumes

- Compute time series of call volumes with 1-minute resolution and compare to baseline (normalized difference)



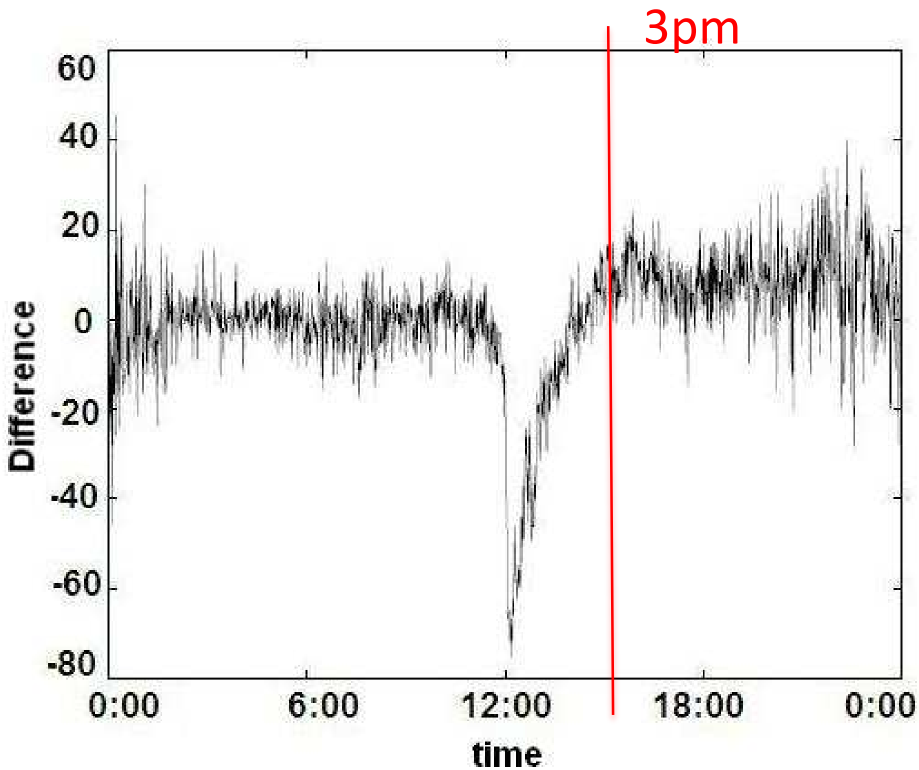
Mexico City



Oaxaca

Call Duration

- Compute time series of call volumes with 1-minute resolution and compare to baseline (normalized difference)

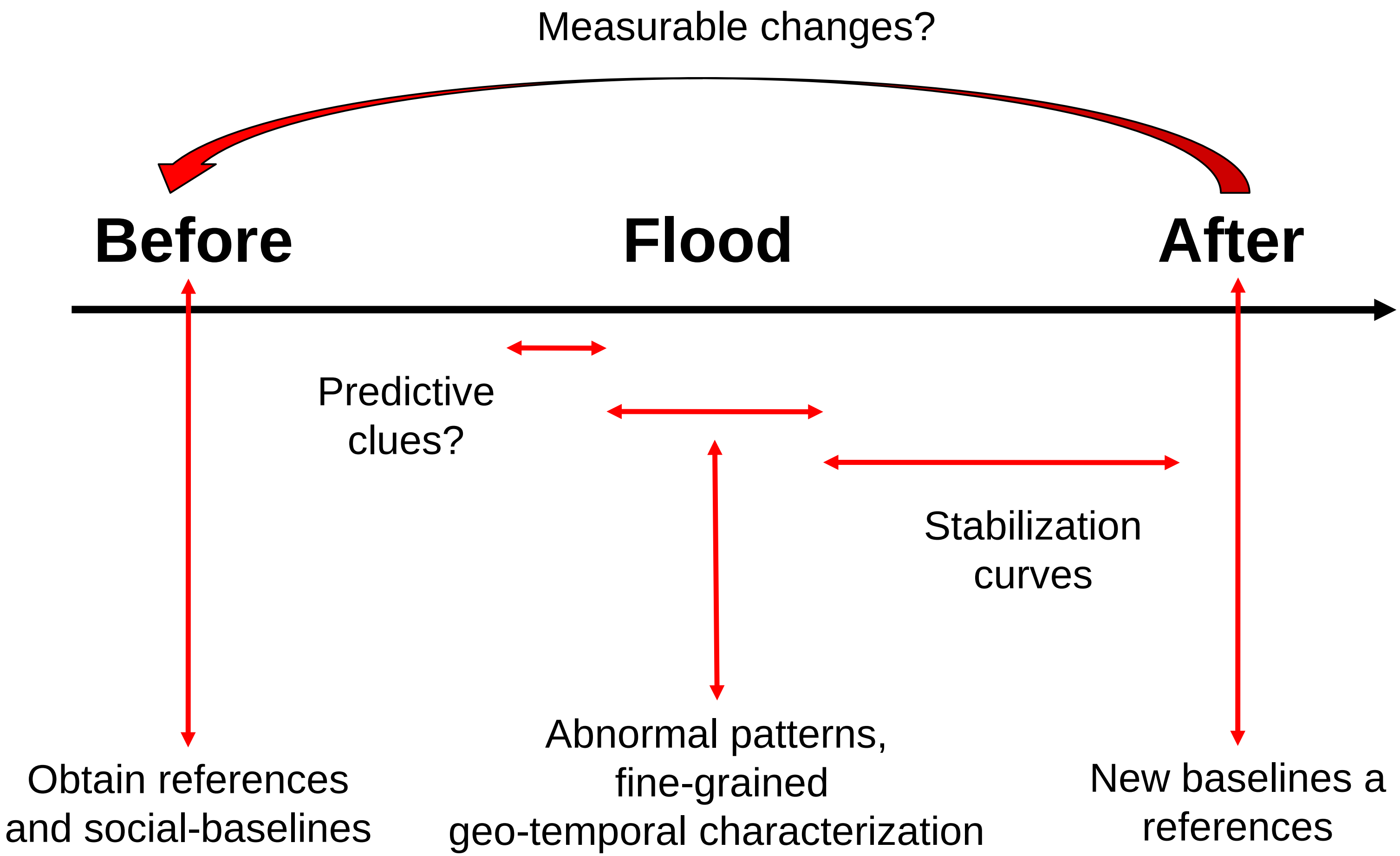


Mexico City

Summary of Hypothesis

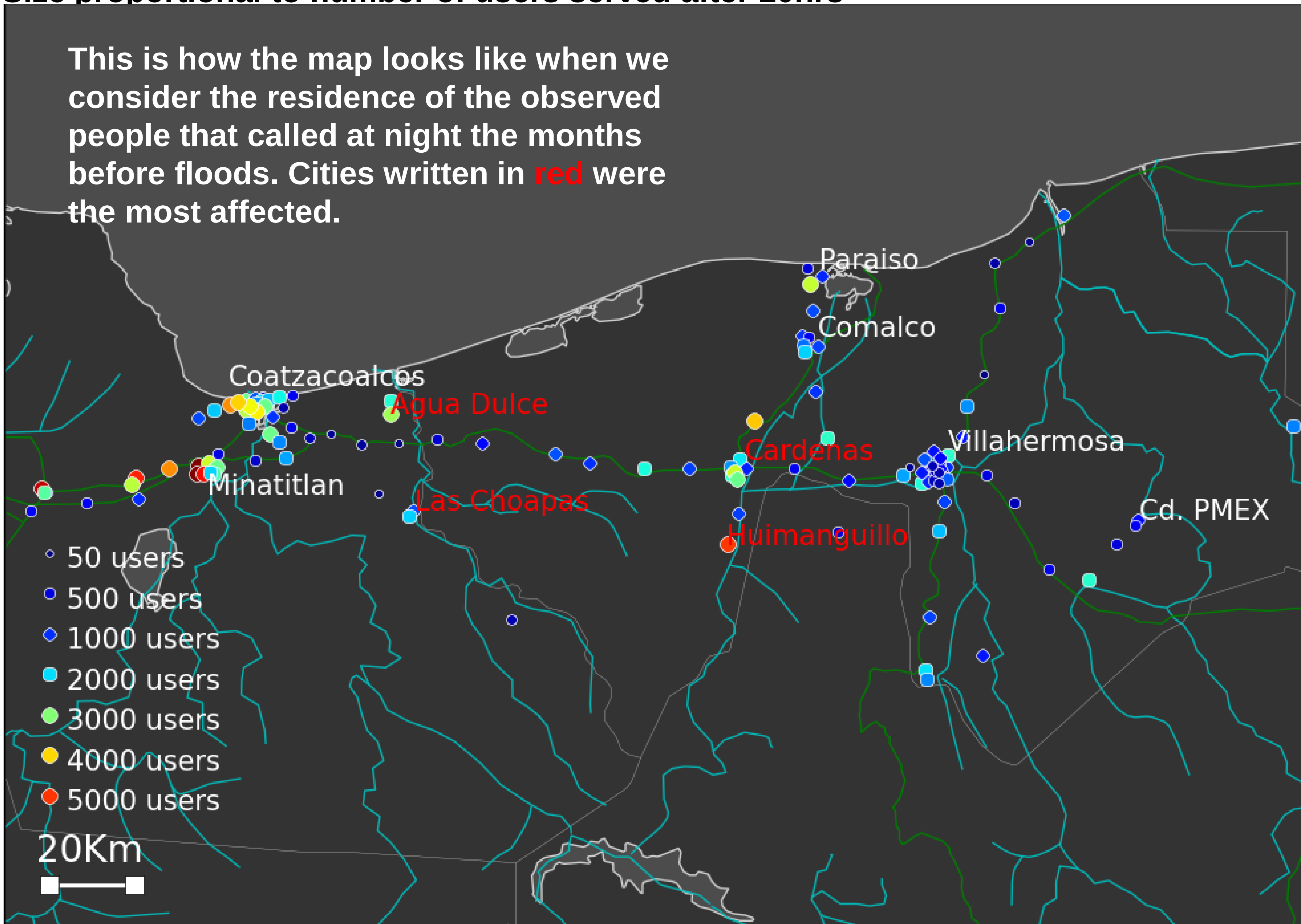
- Larger call volumes right after earthquake
- Shorter call durations ("check call")
- Longer calls at the end of the day
- Highly connected users contact larger number of peers
- Larger mobility patterns during earthquake

Data analysis to study temporal evolution of natural disasters (Tabasco Floods)

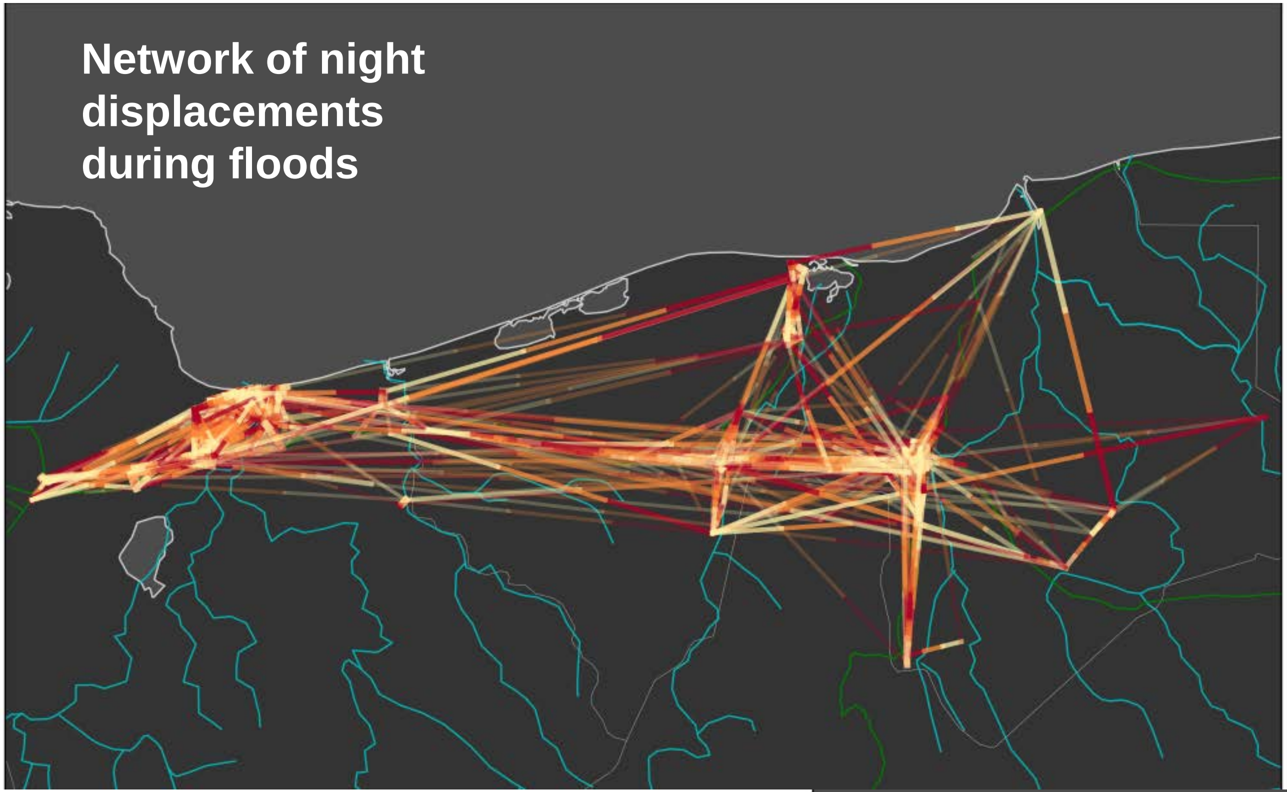


Who lives where

Size proportional to number of users served after 20hrs

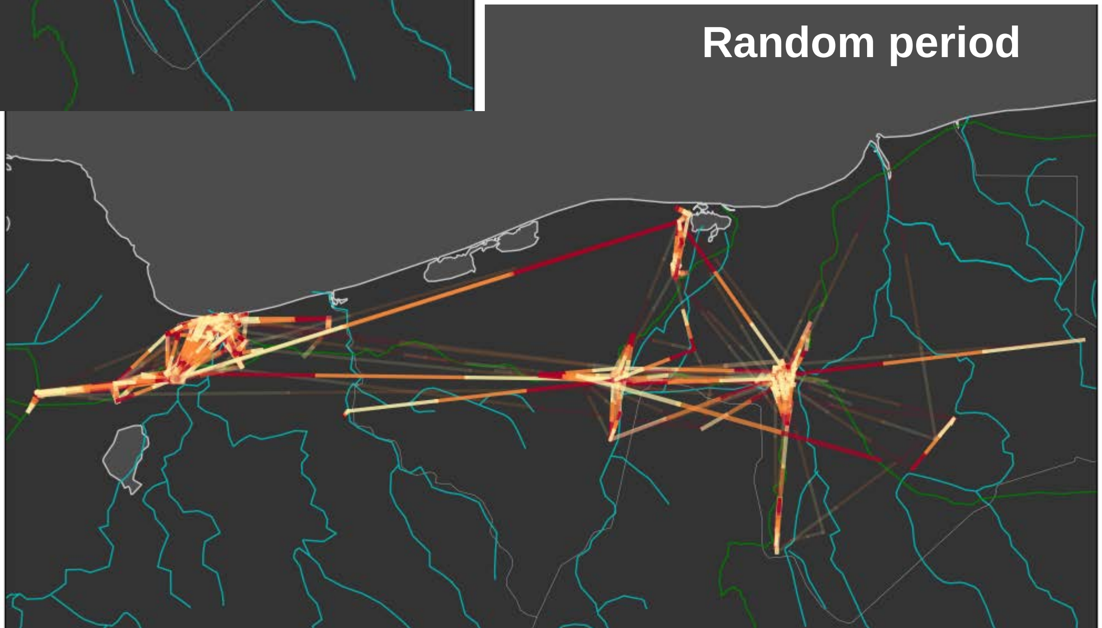


Who changed residence



Here we present a snapshot of the night displacements during floods and the same days one month before. The edge goes from red to pale yellow.

We see that the resulting network is much denser and presents more edges than the previous period.

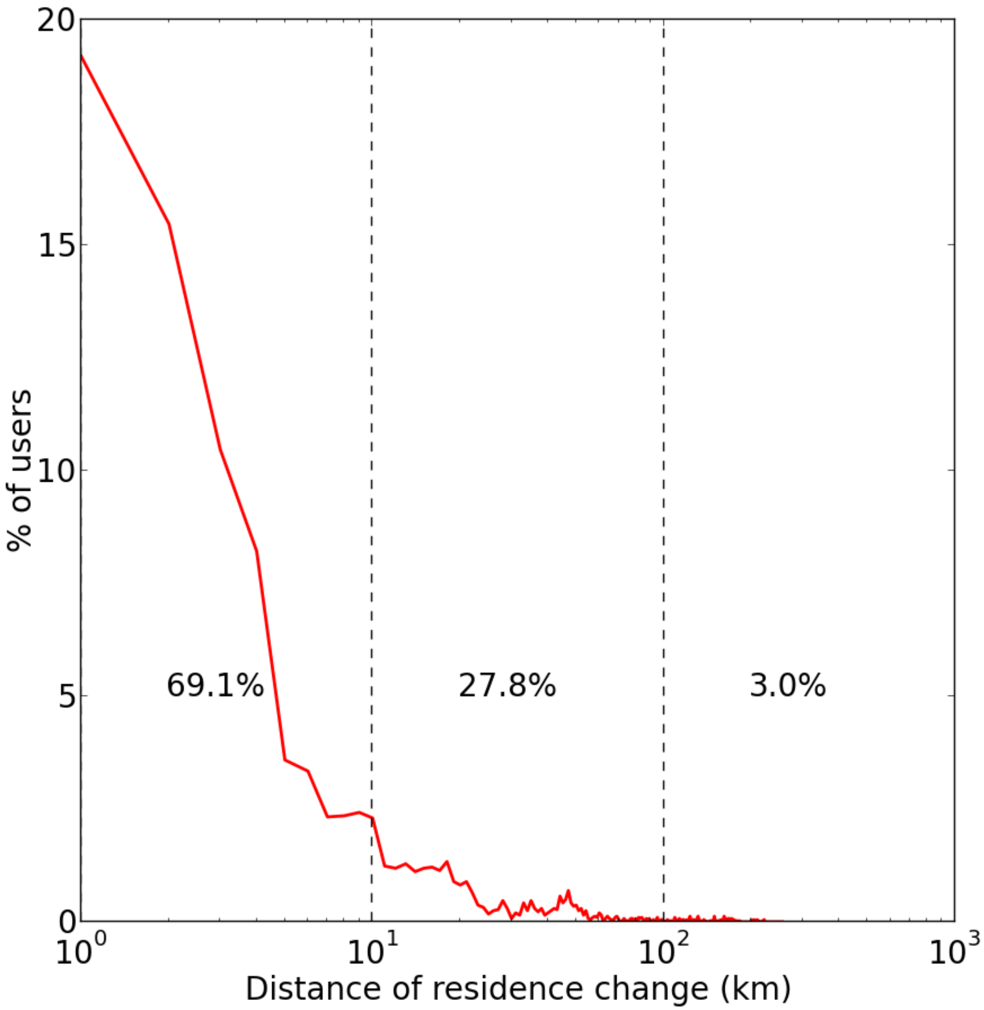


Evacuation centers???

Who changed residence

In this figure we see the distribution of the distances between the old home and the “new” night location.

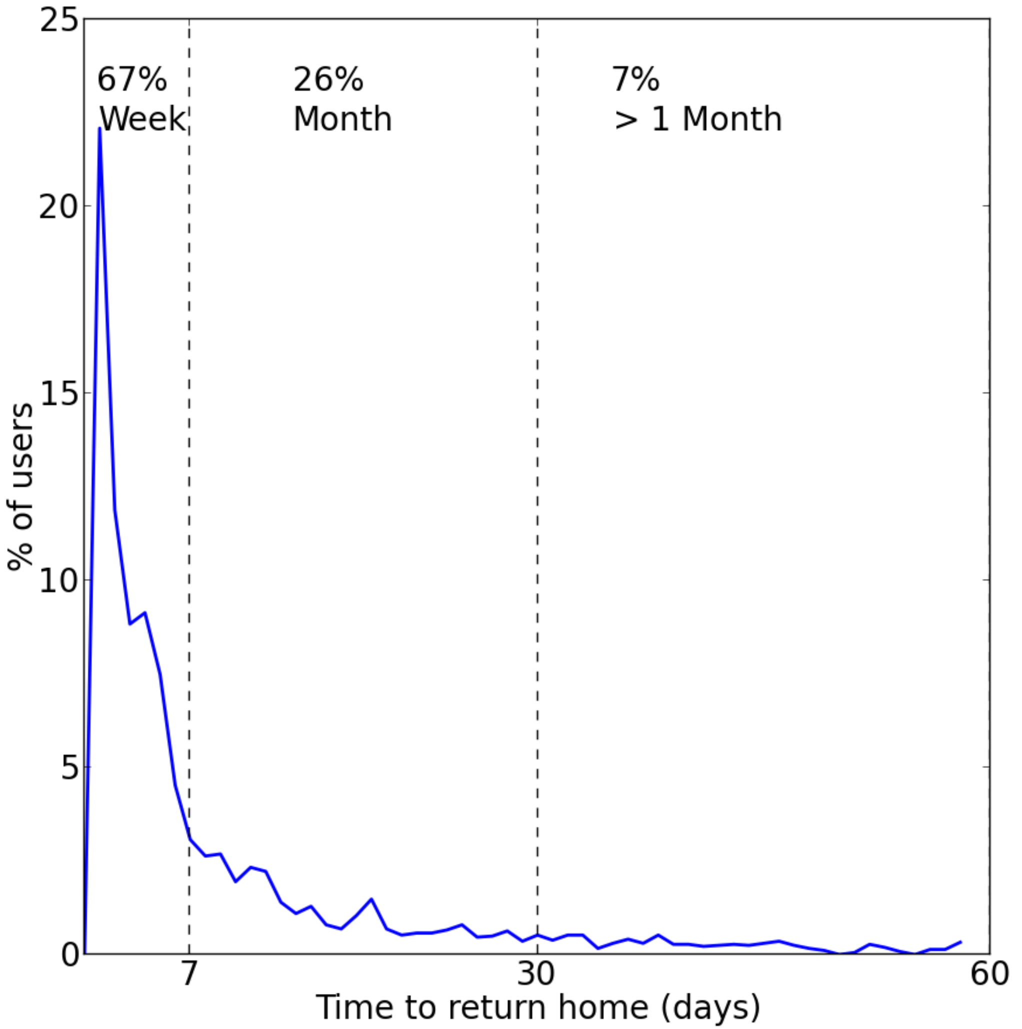
We see that 30% of users displaced more than 10kms and less than 3% of them displaced more than 100kms.



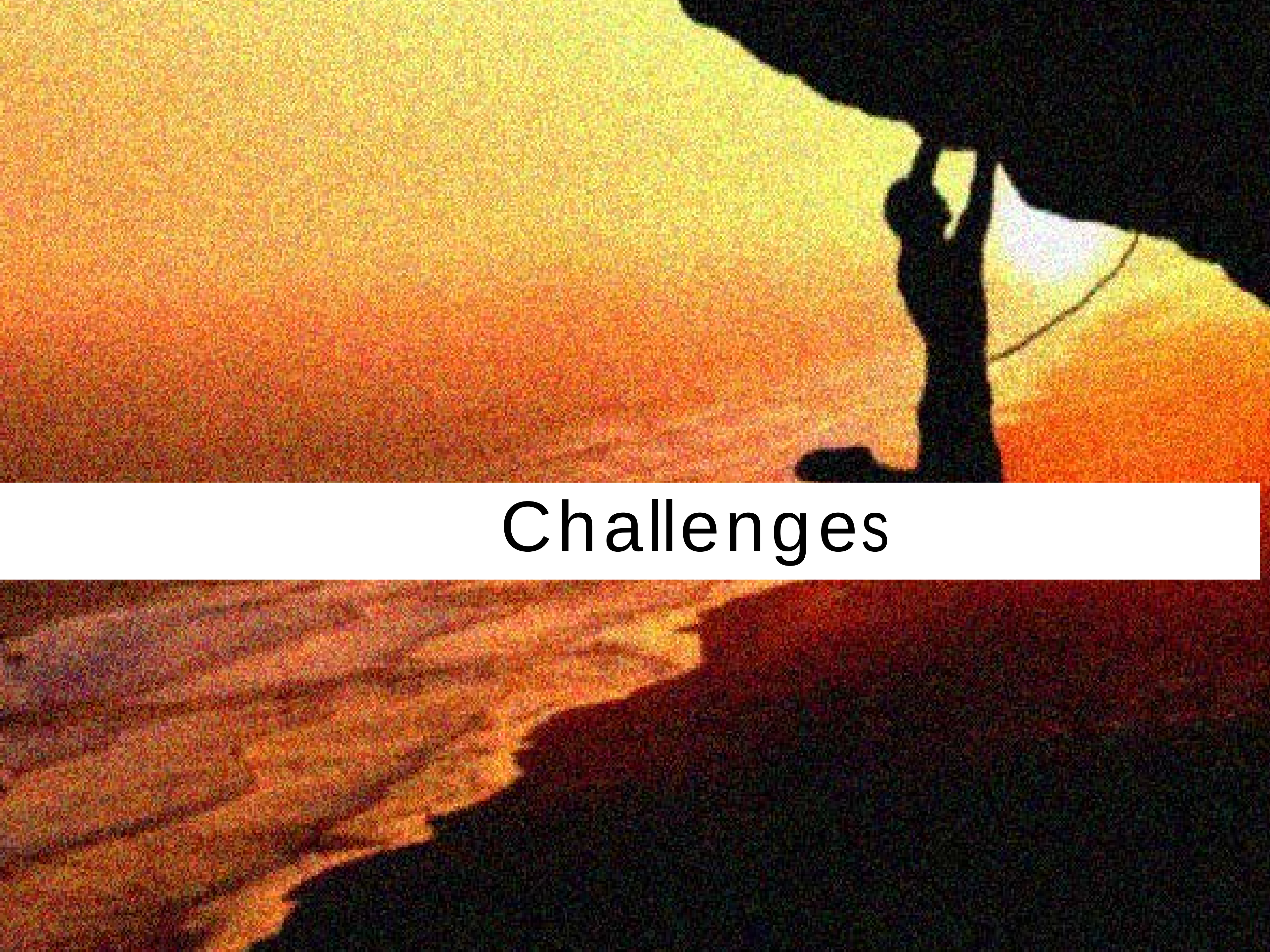
Who returned home

In this figure we see the distribution of the time that took for displaced people to go back home.

We see that more than 60% of users returned in a time lapse of a week, while for 26% of users to between a week and a month o return, more than 8% took more than one month.



Challenges



Ebola and big data

Waiting on hold

Comment (5) | Timekeeper reading list | E-mail | Reprints & permissions | Print

Mobile-phone records would help combat the Ebola epidemic. But getting to look at them has proved hard

Oct 25th 2014 | From the print edition

Rex Features

IN THE battle against Ebola, mobile phones could be invaluable—not just in themselves, as devices that can be used to send people public-health information or let them call helplines, but also because of the data they generate. Phone companies use call-data records, or CDRs, to manage their networks and bill their customers. These records

Regulatory and Social

1. Lack of updated regulation
2. Lack of clear guidelines regarding safe data handling, processing and sharing for humanitarian purposes
3. Risk of potential unintended consequences
4. Risk of creating a digital divide, unbalanced access to data and-or expertise on how to analyze it and make sense of it

Technical

1. Representativeness of the data, generalization
2. Combination of data from multiple sources
3. Real-time analysis and prediction
4. Lack of ground truth → intervention to validate
5. Significant vs substantially significant
6. Correlations vs causality

Privacy and Security

1. Potential privacy risks need to be minimized and understood. Control and transparency
2. Security and traceability of the data
3. Clear code of conduct and ethical principles
4. Strict access control when appropriate

GRACIAS

www.enriquefrias-martinez.info
enrique.friasmartinez AT telefonica.com